

Definiteness in Old Church Slavonic
A Study of How Long and Short Form in
Adjectives Reflect Information Status

Rebecka Lindberg



A thesis presented to The Department of Literature,
Area Studies and European Languages

The University of Oslo

Spring Term 2013

Supervisor: Hanne Martine Eckhoff

Definiteness in Old Church Slavonic

A Study of How Long and Short Form in Adjectives
Reflect Information Status

© Rebecka Lindberg

2013

Definiteness in Old Church Slavonic: A Study of How Long and Short Form in Adjectives
Reflect Information Status

Rebecka Lindberg

<http://www.duo.uio.no/>

Trykk: Reprosentralen, Universitetet i Oslo

Summary

The aim of this master's thesis is to statistically examine how the use of short and long form of attributive and nominalized adjectives and participles in canonical Old Church Slavonic relates to the information status of the NPs concerned. In other words: is the long form of the adjective a marker of definiteness, and does the short form mark the head as indefinite? The study is based on data from the PROIEL¹ corpus, which contains OCS texts that have been annotated for morphology, syntax and information status. Since long form has traditionally been considered a marker of definiteness, the hypothesis is that indefinites should display short form and definites long form. Previous research has claimed that nominalizations show a tendency to use an increasing amount of long form – regardless of the semantics or pragmatics involved – and therefore particular attention will be paid to differences between attributes and nominalizations. The results will show that SF and LF are used based on semantic as well as pragmatic factors and that nominalizations show no erratic behavior.

¹ <http://www.hf.uio.no/ifikk/english/research/projects/proiel/>, http://foni.uio.no:3000/users/sign_in

Acknowledgements

I would naturally like to thank my supervisor, Hanne Martine Eckhoff, who did not only (prior to the writing of this thesis) teach me morphological and syntactic annotation and let me work on the PROIEL corpus, but then also information status annotation and the basics necessary to use the statistical program R. Then she fortunately created the dataset I used, something I would absolutely not have been able to do on my own with my limited technical expertise. Without this dataset, I cannot even begin imagine how I would have written this thesis. She put up with my erratic manner of writing, provided countless tips for improvement, and was the only one who had any idea what I was writing about. Thank you!

Also, thanks to Hilde Spjelkavik Kveseth, who bravely tried to understand the topic and occasionally even remembered it, and to Nicolaas Ervik Groeneboom, who repeatedly encouraged me to be productive, and also helped me understand the basics of statistics. Finally, thanks to Tom-Einar Sørensen, who has been supportive throughout my years as a student at UiO.

Table of Contents

| | |
|--|----|
| Summary | v |
| Acknowledgements | vi |
| 1. Introduction..... | 1 |
| 2. Theory..... | 2 |
| 2.1. What is definiteness? | 2 |
| 2.1.1. Grammatical and semantic definiteness | 2 |
| 2.1.2. Various theories on definiteness..... | 3 |
| 2.1.3. The scope of the definite article and its life cycle | 5 |
| 2.2. Inherent definiteness and overt marking..... | 6 |
| 2.2.1. Demonstratives and definiteness..... | 6 |
| 2.2.2. Possession and definiteness..... | 7 |
| 2.2.3. The case of generics | 8 |
| 2.3. This thesis's take on definiteness | 9 |
| 3. Background..... | 9 |
| 3.1. The pronoun *jъ | 9 |
| 3.2. тъ - definite article or demonstrative? | 13 |
| 3.3. The work of Michael Flier | 15 |
| 3.4. The work of Karen Anne Rondestvedt | 18 |
| 3.5. Remarks on adjectives..... | 20 |
| 3.5.1. Types of adjectives and suffixes | 20 |
| 3.5.2. Comparatives and superlatives | 23 |
| 3.6. Changes in OCS grammar in the 10th and 11th centuries regarding SF/LF | 24 |
| 3.7. Concluding remarks..... | 25 |
| 4. Method | 25 |
| 4.1. Sources | 25 |
| 4.2. PROIEL corpus annotation..... | 27 |
| 4.2.1. NEW and OLD | 29 |
| 4.2.2. ACC-SIT | 29 |
| 4.2.3. ACC-GEN | 29 |
| 4.2.4. NEW-anchored | 30 |
| 4.2.5. ACC-INF..... | 30 |
| 4.2.6. KIND..... | 31 |
| 4.2.7. NON-SPEC..... | 31 |

| | |
|---|----|
| 4.2.8. QUANT | 31 |
| 4.2.9. Challenges related to the transfer of IS-annotation from Greek | 32 |
| 4.2.10. Concluding remarks regarding IS-tags..... | 33 |
| 4.3. Treatment of data | 33 |
| 4.3.1. Data processing | 33 |
| 4.3.2. A first view of the IS/strength distribution..... | 34 |
| 4.3.3. Exclusions | 34 |
| 4.3.3.1. Non-varying suffixes and lemmas | 34 |
| 4.3.3.2. The quantifiers <i>тъногъ</i> and <i>въсѣкъ</i> | 34 |
| 4.3.3.3. The ordinal numeral <i>drugъ</i> | 35 |
| 4.3.3.4. The present participle <i>sy</i> | 35 |
| 4.3.3.5. Coordinated attributes | 36 |
| 4.3.3.6. Predicative expressions | 36 |
| 4.3.3.7. Conjunct participles..... | 37 |
| 4.3.3.8. Absolute datives | 37 |
| 4.3.3.9. Cooccurrences with demonstratives..... | 37 |
| 4.3.3.10. Cooccurrences with possessives | 39 |
| 4.4. Concluding remarks..... | 41 |
| 5. Analysis – Attributes, nominalizations, appositions and vocatives..... | 41 |
| 5.1. Attributes..... | 43 |
| 5.1.1. The -ъsk suffix..... | 43 |
| 5.1.1.1. Adjectives formed from unique entities | 44 |
| 5.1.1.2. Miscellaneous..... | 45 |
| 5.1.1.3. Adjectives formed from places or peoples..... | 46 |
| 5.1.1.4. Conclusion regarding -ъsk adjectives | 48 |
| 5.1.2. Short and long form in the NEW category for attributes..... | 48 |
| 5.1.2.1. IS-tag NEW and short form attribute | 49 |
| 5.1.2.2. IS-tag NEW and long form attribute | 49 |
| 5.1.2.3. Short and long form for NEW-anchored referents..... | 52 |
| 5.1.3. Short and long form in the OLD category for attributes | 54 |
| 5.1.3.1. IS-tag OLD and long form attribute | 54 |
| 5.1.3.2. IS-tag OLD and short form attribute..... | 55 |
| 5.1.4. Short and long form in the ACC-category for attributes | 59 |
| 5.1.4.1. IS-tag ACC-GEN and long form attribute | 59 |

| | |
|---|-----|
| 5.1.4.2. IS-tag ACC-GEN and short form attribute..... | 60 |
| 5.1.4.3. Strength in the ACC-INF and ACC-SIT categories for attributes | 62 |
| 5.1.5. Short and long form in the NON-SPEC category for attributes | 64 |
| 5.1.6. Short and long form in the KIND category for attributes | 68 |
| 5.1.7. Attributes to proper nouns | 70 |
| 5.1.8. Word order for attributes | 72 |
| 5.1.9. Conclusion regarding attributes | 73 |
| 5.2. Nominalizations | 73 |
| 5.2.1. Short and long form in the NEW category for nominalizations | 74 |
| 5.2.1.1. IS-tag NEW and short form nominalization..... | 74 |
| 5.2.1.2. IS-tag NEW and long form nominalization | 75 |
| 5.2.1.3. Short and long form for NEW-anchored referents..... | 77 |
| 5.2.2. Short and long form in the OLD category for nominalizations | 79 |
| 5.2.2.1. IS-tag OLD and long form nominalization | 79 |
| 5.2.2.2. IS-tag OLD and short form nominalization | 80 |
| 5.2.3. Short and long form in the ACC category for nominalizations..... | 82 |
| 5.2.3.1. Strength in the ACC-GEN category for nominalizations | 82 |
| 5.2.3.2. Strength in the ACC-INF category for nominalizations..... | 83 |
| 5.2.3.3. Strength in the ACC-SIT category for nominalizations | 84 |
| 5.2.4. Short and long form in the NON-SPEC category for nominalizations | 85 |
| 5.2.4.1. NON-SPEC nominalized participles..... | 85 |
| 5.2.4.2. NON-SPEC nominalized adjectives | 86 |
| 5.2.5. Short and long form in the KIND category for nominalizations | 89 |
| 5.2.6. Short and long form in the QUANT category for nominalizations | 91 |
| 5.2.7. Conclusion regarding nominalizations | 93 |
| 5.3. Appositions | 94 |
| 5.3.1. Nominalized appositions on proper nouns | 94 |
| 5.3.2. Attributes to appositions..... | 96 |
| 5.3.3. Nominalized appositions on common nouns..... | 98 |
| 5.3.4. Conclusion regarding appositions | 99 |
| 5.4. Vocatives | 100 |
| 5.4.1. Nominalized vocatives..... | 101 |
| 5.4.2. Attributive vocatives | 102 |
| 5.4.3. Conclusion regarding vocatives..... | 102 |

| | |
|---|-----|
| 6. Concluding remarks..... | 103 |
| 6.1. Summary and findings..... | 103 |
| 6.2. Evaluation of the procedure..... | 105 |
| 6.3. Further research..... | 105 |
| Works Cited | 106 |
| List of tables | 109 |
| Appendix 1 - Change of IS-tag category | 110 |
| Attributes..... | 110 |
| Nominalizations..... | 112 |
| Appendix 2 – Additional data | 115 |
| Examples from chapter 5.1.2.3, Attributive NEW-anchored that are erroneously anchored for various reasons..... | 115 |
| Examples from chapter 5.2.1.3, Nominalized NEW-anchored that are erroneously anchored for various reasons..... | 116 |
| Examples from chapter 5.2.2.2, OLD SF that should display LF | 117 |
| Division of adjectives formed with the -ьsk suffix from chapter 5.1.1. | 117 |
| Abstract/Non-abstract division from chapter 5.2.4.2. | 118 |

1. Introduction

Old Church Slavonic (OCS) is the first literary language of the Slavs and was created in the 9th century for the translation of biblical texts. It is traditionally believed to have possessed a limited system of overt definiteness marking, relying on the inflection of adjectives and participles. The bare form of the adjective is considered indefinite, whereas the long form, combining the bare form with the pronoun **jb*, individualizes the head and renders it definite. The goal of this thesis is to examine, with the help of information status tagging, how long and short form attributive and nominalized adjectives and participles in OCS relate to the pragmatic roles the NPs play. Put more simply: do short and long forms on adjectives in OCS express definiteness? This distinction does not exist in modern Slavic languages, and the question is then whether the system is disappearing already in the canonical texts, or if it is still functional. Does the behaviour of nominalizations show the beginning of the end for the long form as a marker of definiteness, as some previous research has indicated? Are generics marked in some uniform manner?

The texts used are the Gospels as well as some later texts, spanning the 9th and 11th centuries. These texts are translations from Greek, which makes plentiful use of definite articles on both nouns and adjectives. The original Greek Gospels are also present in the PROIEL corpus, making it possible to see how the OCS translators have chosen to convey arthrous and anarthrous Greek NPs using short and long form adjectives respectively. Previous research has been carried out on the same material but has left certain categories of words without any satisfying explanation as to their behavior. With the help of the information status analysis some of these issues may be resolved.

The results will show that short and long forms as a system of definiteness marking is still very much alive in the 9th to 11th century texts. Short form on adjectives signals that something is new or non-referential, whereas long form marks referents as old or generally known. It will also be demonstrated that nominalizations do not behave differently from attributes, that appositions and vocatives show short and long form when expected to do so, and that generics show long form. Word order will be shown to be of little importance in the system of definiteness marking.

The structure of this thesis is as follows: In chapter 2, theories on definiteness will be presented: the concepts of grammatical and semantic definiteness, inherently definite environments as well as the special case of generics. The formation of the long form is then explained in chapter 3, together with comments on OCS demonstratives and previous research into the SF/LF dichotomy, the nature of adjectives in OCS and relevant changes in OCS grammar in the time period at hand. In chapter 4, the PROIEL corpus annotation is explained. Since it is of interest in the examination of the attributive

adjectives and participles to deal with as “pure” modifiers as possible, where hopefully nothing but pragmatic factors affect the distribution of short/long form, a great deal of material will then be excluded previous to the final analysis. These exclusions are also found in chapter 4. The bulk of the analysis is found in chapter 5, first separated into attributive and nominalized adjectives/participles. In each of these groups, short/long form occurrences are examined according to the information status of the referents. After that, appositions are dealt with apart from nominalizations to see if they show any particular behavior, and finally vocatives are treated separately altogether.

Throughout the thesis, examples are given in OCS and transliterated NT Greek, English and occasionally other European languages² where this is of interest. Due to the narrowing down of the final dataset used, most of the pertinent examples are presented in the text. However, when something of interest has been left out due to space restraints, this information can be found in Appendix 2. The findings of this thesis are presented in the conclusion in chapter 6.

2. Theory

The theory chapter will deal with concepts that are crucial for this thesis, notably the ideas of definiteness and genericity. The theories of what makes something definite and overtly marked as such in various languages will then be tied to the different IS-tags used in the PROIEL corpus in chapter 4.2.

2.1. What is definiteness?

What does it mean to say that something is definite? In an examination of phenomena related to definiteness, it is crucial to have an idea of what definiteness actually is. This is a question that belongs both in the field of linguistics and in that of philosophy, and there is no lack of theories describing what definiteness is. Here an attempt at a synthesis of various theories will be presented, mainly based on Christopher Lyons’s *Definiteness* (1999), where Lyons argues that definiteness is a grammatical category, related to semantic and pragmatic concepts.

2.1.1. Grammatical and semantic definiteness

Grammatical and semantic definiteness are concepts that are important not to confound, and even though the *semantic* feature [±DEF] is present in all languages in one way or another, not all of them express it *grammatically* through article systems (which are, furthermore, not the only strategy used in languages to convey definiteness). All languages have demonstratives and personal pronouns,

² Bible quotes are all taken from <http://www.biblegateway.com>, OCS and NT Greek examples are taken directly from the PROIEL corpus.

which are, in Lyons's (1999, p. 48) words "perhaps inherently definite". Individual words are also – regardless of context – inherently definite, which is explained by their uniqueness. There is, within a certain cultural context, typically only one *sun*, *Heaven*, *God*, etc. Even for this type of referent, article use varies in languages that use them. In a similar vein, proper nouns are unique in that they always refer to only one person. Even though there may be very many Roberts in the world, every typical use of "Robert" will refer to one single Robert. Some languages mark proper nouns with definite articles (thus overtly signaling that this is something definite), such as Ancient and Modern Greek, whereas other languages leave them unmarked. Therefore, in languages that make use of grammatical definiteness *semantic definiteness* and *grammatical definiteness* need not necessarily overlap. When it comes to OCS, the [+DEF] feature is believed to be signaled by LF³ in the attribute, meaning that the only nouns⁴ – other than those accompanied by demonstratives – that can be explicitly, grammatically, marked for definiteness are the ones modified by certain types of adjectives (Flier, 1974, p. 67).

2.1.2. Various theories on definiteness

There are several different theories trying to define what "definite" really means. Here we can, on the one hand, mention the theories of *familiarity*, *association*, *identifiability*, and, on the other, *inclusiveness* and *uniqueness*. Among these, identifiability and inclusiveness are usually accepted as being the most important explanations as these two can be argued to encompass the other notions.

The familiarity theory corresponds to what usually comes first to mind when discussing the problem of definiteness: what is known to both speaker and hearer. On this account, "the" signals mutual familiarity with the referent. However, it is not very difficult to find examples of things that are certainly not familiar to both speaker and hearer, but that are still accompanied by a definite article. This shows that familiarity alone cannot explain all uses of definiteness marking. Thereby, we pass into the field of *association* and *identifiability*. For example, the mention of a taxi conjures up other things that are associated with taxis, and you can say "I paid *the driver* in cash", without the addressee knowing who this driver actually is (Lyons, 1999, p. 4). The referent is identifiable by association. By using "the", the speaker signals that the addressee should be able to *identify* the object in question, but not necessarily because he/she is familiar with it. "So while on the familiarity account *the* tells the hearer that he knows which, on the identifiability account it tells him that he

³ Throughout the thesis, Long Form and Short Form will be abbreviated as LF and SF respectively.

⁴ Inherently unique NPs without attribute remain unmarked, unless they are animate and appear as direct objects. As direct objects, animate masculine nouns (unique ones in particular) more frequently have an accusative case form identic to the genitive (Eckhoff, Animacy and differential object marking in Old Church Slavonic, 2011, pp. 2, 19), whereas other nouns display either an accusative form identical to the nominative or a specific accusative form (the feminine -u). As such, there *are* other strategies besides the use of regular adjectives present in OCS to mark definiteness.

knows or can work out which" (Lyons, 1999, p. 5). As Lyons (1999, p. 8) points out, in cases of association the hearer will not be able to *actually* identify the referent, that is, know who that person is and later pick him/her out in a crowd. The referent is identifiable only within the context, relative to which it is *unique*.

The idea of uniqueness is an important one. "Can you get me *the towel*?" suggests that there is one unique towel in that same room, within view, and the article use will make the addressee look around for that towel. "Can you get me *a towel*?" just suggests that there *are* towels (not necessarily within view), and the speaker wants one of them. Lyons (Ibid., p.8) differs between the just mentioned *relative uniqueness* (which is situational) and *absolute uniqueness*, dealing with inherently unique referents (the sun, Satan, etc.), to explain why certain NPs get a definite article. As we move from the singular to the plural, the idea of uniqueness can be subsumed under, or assimilated to, that of inclusiveness. Lyons argues that "all" is the plural equivalent of "the". You can include a certain object into a predefined mass of objects, but a single object can also be such a definite unity all on its own: "When the noun phrase is singular, inclusiveness turns out to be the same as uniqueness, because the totality of the objects satisfying the description is just one" (Ibid., p. 11). As examples, we can take two of Lyons's sentences:

- (1) **The man** who comes with me will not regret it.
- (2) We're looking for **the vandals** who broke into the office yesterday.

There is only one man who will match the description in (1), i.e. he is unique (but not previously known, familiar or identifiable by association). The article in (2) signals that "we" are looking for *all* the vandals who match the description of having broken into the office, i.e. the definite article is inclusive. The referents are unique, but supposedly not known to anyone.

Extremes (superlatives) and polar opposites⁵ (left/right) are also unique, either relative to a referent or in an absolute manner. A point of interest is that languages that do mark definiteness grammatically always do so for superlatives⁶ (Lyons, 1999, p. 246).

To simplify matters, these accounts of definiteness can be summarized as *identifiability* and *inclusiveness*. We end up with four sorts of possible uses for the definite article: *anaphoric*, which operates within the linguistic context, referring to things that are familiar to both hearer and speaker since they have already been mentioned; *situational*, in which case the hearer is supposed to figure

⁵ These polar opposites are referred to by Dahl (2004, p. 153) as "selectors" and he says that "All these categories share a common semantics – they are all 'inherently definite' in that the noun phrases they are used in normally have definite reference by virtue of their meaning."

⁶ As Lyons (1999, p. 246) further points out, at least in English (but this is also true for Scandinavian languages), the only other possibility with superlatives is possessive pronouns, as in "my cleverest student".

out or identify what is referred to in a given situation; *general knowledge*, which deals with unique entities known to all and also generic referents, and finally the *associative* use that can be combined with the other three (Lyons, 1999, p. 158). Not all of these need to be encoded similarly in all languages, and this can be summarized by the following quote from Lyons (1999, p. 159) on semantic and grammatical definiteness:

For languages in general there is a range of noun phrase uses which can in principle be characterized as definite, because they can be described in terms of identifiability or inclusiveness. These uses represent “semantic definiteness”, but this is not what articles encode. A given language need not treat the full range of these uses as grammatically definite; so the feature specification [+ Def] can segment the semantic field at different points in different languages, its range in a particular language being shown by which uses require the presence of a definite article or other definiteness marker.

Lyons (1999, pp. 277-278) considers definiteness to be a grammatical phenomenon on par with tense or number. Grammaticality does not always match up with semantics, since singular items can be pluralia tantum in form (trousers, scissors), tenses and moods can be used in ways that seemingly break the rules but that are conditioned by grammatical features (he cites uses of *le subjonctif* in French where indicative is used in other languages). In the same way, grammatical definiteness is not always 100% aligned with the semantics underlying it, and does not find perfect correspondence cross-linguistically. Generics, which will be discussed in 2.2.3, are a good example of this, as they show use of the definite article in some languages, but not in others.

2.1.3. The scope of the definite article and its life cycle

Languages differ in the range of NPs they explicitly mark as definite using articles. Lyons (1999, p. 337) presents a simplified account of how the article is used in different languages in the following list:

- 1 (English): simple definite
- 2 (French): simple definite, generic
- 3 (Italian): simple definite, generic, possessive
- 4 (Greek): simple definite, generic, possessive, proper noun

One of the goals of this thesis is to determine how a similar list would look for OCS. The hypothesis, based on work with OCS texts and previous research, is that the OCS LF covers several categories of definite referents, and would end up with a distribution similar to that of Italian/Greek, with the important distinction that only nouns modified by attributes can carry this overt marking. Considering the posited scope of the use of definiteness marking, it is perhaps all the more interesting that most Slavic languages today lack any such grammatical marking. This is not, however, an unusual phenomenon: definiteness marking does arise to then disappear.

Greenberg (1978) argues that definite articles have a life cycle consisting of three stages that normally ends in the article becoming a marker of nominality. Thus, definite articles often start out as demonstratives or personal pronouns, develop into markers of definiteness of various scope, and then turn into gender affixes, class markers, etc. Keenan (2011, p. 2), in a conference contribution entitled “Greenberg Revisited: Diachronic development of article systems & the structure of DP”, dealing with the development of article systems, states that:

Once a language has begun – for whatever reason – to mark discourse reference overtly in the functional structure, its use of articles in this function will gradually spread in a predictable manner to a broader (and progressively less referentially ‘strong’) range of noun phrases.

This more or less sums up the life cycle of the marker of definiteness that the LF addition **jb* (presented in detail in chapter 3.1) to the adjective is in OCS. It starts out as a pronoun/demonstrative, spreads as a marker of definiteness to an array of different categories of referents (including generics), and ends up as a marker of attributive/appositive adjectives and participles. The Slavic languages that still have LF adjectives today use it mainly in the attributive/predicative split. This split is not “perfect”, as LF exists in the predicative position as well as in the attributive, and there are remnants of the definite/indefinite division left in some languages.

2.2. Inherent definiteness and overt marking

2.2.1. Demonstratives and definiteness

Demonstratives, since they refer to something that is identifiable to the addressee, are inherently definite (Lyons, 1999, p. 20). The use of a demonstrative signals that the speaker expects the addressee to understand what item is being referred to, that is, it is a question of actual identifiability (Lyons, 1999, p. 18). When someone says “Pass me *that book*”, the deictic demonstrative signals that there should be a suitable book within reach that the addressee will be able to properly identify. What would then happen in OCS if the phrase was “Pass me *that big book*”? In OCS, one could believe that, whenever there is already a demonstrative present, no LF on the attribute would be needed to mark the NP as definite, as is indeed the case with nouns that stand alone without an adjectival modifier. This is a question of either being economical in the use of overt marking, or explicit. Michael Flier often refers to double marking of definiteness, as well as marking of inherent definiteness, as “redundant marking”. It is obvious that economy in definiteness marking is by no means a necessary trait among world languages. Both Greek and Scandinavian languages use seemingly unnecessary definiteness marking – from a typical Anglo-centric point of view where one

marker would be deemed sufficient. Speakers of these languages obviously see a need to mark definiteness very explicitly, just like adjectives in some languages need to be adjusted according to gender or number or both, even though this should by the same argument already be clear from article use or the head word itself. Based on previous research, LF is expected in the presence of demonstratives in OCS, meaning that explicitness supersedes economy.

2.2.2. Possession and definiteness

Possession is another category that is important for the discussion on definiteness. In examining OCS attributes, we will pay close attention to attributes that cooccur with possessive modifiers. As opposed to demonstrativeness, possession does not necessarily entail definiteness (Lyons, 1999, p. 24). Whether possessives entail definiteness or not depends on the nature of the constructions present in the language. Lyons differs between two types of languages, AG (adjectival-genitive) and DG (determiner-genitive) languages. In AG languages, possessives are adjectival, whereas they occupy the position of the definite article in DG languages. As Lyons points out, both types of constructions are often present in the same language. English is a DG language, where the possessive automatically introduces definiteness (Ibid., p. 24). With AG languages, there must be a definite article present with the possessive in order for it to be definite (Ibid., p. 130). A typical example of an AG language is Classical Greek, where the genitive case is used to mark possession, and where *pronoun possessives* (**the my** book) behave the same way as *full noun phrase possessors* (**the child's** book) (Ibid., p. 134). OCS uses possessive adjectives, genitive and dative constructions, but also possessive pronouns to mark possession. The various possessive constructions will be checked for occurrences of SF/LF adjectives in chapter 4.3.3.10. An example of such a construction can be found in (3)⁷:

- (3) се есть тѣло мое **даемое** за вы (CM, Luke 22:19, 41359)
 touto estin to sōma mou **to** hyper humōn **didomenon** (21629)
 This is my body, **which is given** for you.

In the example above, the participle modifying *tělo*, “body”, *dajemoje*, “which is given” is in the LF and there is a possessive pronoun present that clearly individualizes the noun. It is seemingly the same pattern as that expected for demonstratives; the possessive does not render LF redundant, and OCS seems to behave like an AG language. What makes DG type constructions inherently definite is, according to Lyons (1999, p. 134), not the possession in itself, but rather the position in the NP, i.e. where the definite article is usually found. The AG constructions, on the other hand, place the

⁷ In examples, the head word (when there is one) is most often underlined, whereas the attribute is in bold. Underlined and bolded words in the various translations correspond to each other. Following the OCS text, the manuscript it comes from is listed (CM or CS), followed by the text in question and the sentence ID by which it can be accessed in the PROIEL corpus (see Appendix 1).

possessive deeper within the NP, leaving the determiner position free to be filled by an article or a demonstrative. Furthermore, Lyons (Ibid., p. 132) suggests that DG constructions are most likely restricted to languages that have definite articles, whereas AG constructions are not.

For both demonstrative and possessive pronouns cooccurring with attributes, Rondestvedt (1986, p. 120) claims that long form is used exclusively.

2.2.3. The case of generics

As Pelletier (2010, p. 9) points out “much of our common sense knowledge of the world is expressed by [...] generic sentences”, making generics an important category of expression. For this category it is important to separate between generic sentences (such as (4)) and generic referents (such as “the rich”). Since the same phrase types that are used for generics are also used for other types of referents, Lyons (1999, p. 179) suggests that genericity is not a “primitive category of semantic or syntactic description”. Definites, indefinites and bare plurals can be used for generics, depending on the language, and some languages also make use of several different types of NPs, as does English. The following examples from Lyons (1999, p. 179) and Rondestvedt (1986, p. 79) can be cited:

- (4) a. A dog has four legs.
 b. The dog has four legs.
 c. Dogs have four legs.
- (5) Mice were chasing my cat all over the house.

Here we can first of all note that the bare plural in (5) is not at all generic, but indefinite specific, whereas all of the examples in (4) with varying grammatical expression refer to dogs in general. Generic statements do not have to include all members of a class to be considered true, as in “The lion has a mane” since not *all* lions have manes. As such, generic sentences are “exception tolerant” (Pelletier, 2010, p. 9).

Lyons (1999, p. 198) concludes his discussion on generics by stating that they are semantically, although not always grammatically, definite. This can possibly be explained by their inclusiveness, since they include more or less all members of a class, and this would be especially true of plural generics. But, he also says “They can perhaps also be said to be familiar, and therefore to meet the criterion of identifiability” (Ibid., p. 197). Generics are an important group, but as can be seen from the examples above, a group that poses certain difficulties when it comes to defining its grammatical expression. Since markers of definiteness play an important role in this group, they are of interest to look into for OCS as well.

2.3. This thesis's take on definiteness

This thesis will focus on morphologically expressed definiteness as an expression of identifiability, subsuming things that are “familiar” (i.e. already known) and therefore identifiable, and “accessible” (i.e. inferable from something else and/or unique relative to the context) to speaker and addressee and therefore identifiable. Referents that are generally known and often unique fall under the idea of familiarity, as do those that have already been introduced to the addressee and are referred to in subsequent mention. NPs that can easily be identified by the addressee due to situational circumstances belong in the “accessible” camp. There is something to be said for the idea of inclusiveness as well, which explains definiteness when an entire class of referents is envisioned. Quantified referents are expected to show LF for quantifiers carrying the nuance of “all” or “every” based on inclusiveness, but in the case of “many”, there should be no such constraint.

3. Background

This chapter will deal with previous research related to SF/LF carried out by in particular Flier (1974), Dimitrova-Vulchanova and Vulchanov (2009), as well as Rondestvedt (1986). They have all looked at the phenomenon of SF/LF on adjectives and participles in relation to definiteness, but without the addition of a systematic information status analysis. They have also partly used different sources, in particular Rondestvedt, who works with original Slavic texts and Dimitrova-Vulchanova and Vulchanov who have younger, Bulgarian texts as well as overlap with texts present in the PROIEL corpus. First and foremost, the origin of the LF marker, the **jъ* pronoun/demonstrative, will be examined.

3.1. The pronoun **jъ*

This sub-chapter will present some general comments on the pronoun **jъ*, the demonstrative pronouns used in OCS and the supposed place of **jъ* among these, as well as the full declensions of both the personal pronoun and the SF/LF versions of OCS adjectives and participles.

The pronoun or morpheme **jъ* has been referred to as a demonstrative pronoun in the literature on OCS by some linguists (Lunt, Flier, Čaburgaev), but whether it should actually be considered a demonstrative or not is questionable. At least by the time of the Gospel texts, it seemingly fills no such function. Instead, it is used as a personal pronoun in all cases except the nominative⁸, and as a LF morpheme on adjectives and participles. Rondestvedt (1986, p. 1) refrains from clearly placing it among the demonstratives, saying that the *morpheme* **jъ* is “derived from the Indo-European

⁸ This pronoun never appears in the nominative, hence the asterisk. In the nominative, *тъ* or *онъ* are used, whereas **jъ* appears in the oblique cases (Lunt, 2001, p. 63).

demonstrative *ei- or the relative *io-” (my emphasis). Flier (1974, p. 67) presents the same history, but says that the two IE pronouns merged in the OCS pronoun. Because of the semantic closeness of demonstratives and definite articles, and the claims made by Dimitrova-Vulchanova & Vulchanov regarding LF/articles that will be discussed farther on, the different OCS demonstratives will be presented here with examples:

- (6) къто **съ** естъ ѿко вѣтри и море послоушаѣтъ его (CM, Matt 8:27, 50738)
 ‘What sort of man is **this**, that even the winds and the sea obey him?’
 Literally “who *this* is, that...”, cf. Greek *potamos estin houtos*, “from what country/of what sort is this (one)”.
- (7) и исцѣлѣ отрокъ его во **тъ** часъ (CM, Matt 8:13, 50719)
 kai iathē ho pais en tē hōra **ekeinē** (14923)
 And the servant was healed in **that** hour.
- (8) егуптъстии зноѣве оубо. ничимъже сѣтъ хоуждѣши. вавулоньскыа **оноа**
 пештъница. (CS, Paul the Simple, 58719)
 For the Egyptian heats are in no way worse than **the** Babylonian furnace.⁹
 Literally “Babylonian that furnace” in the genitive.

The demonstratives seen here are: *съ* (this), *тъ* (that), *онъ* (that over there), in increasing degree of distance from the speaker¹⁰. *Онъ* used as a demonstrative is mostly found in the set expression *онъ поѣ*, meaning “the other side”, and otherwise as a personal pronoun.

Finally, there is **ѣ* (he, this) – if we accept that **ѣ* has at some point been a demonstrative – which stands out from the rest in being the form that carries the least deictic content (Flier, 1974, p. 61). Since **ѣ* does not indicate closeness or remoteness to the speaker it is suitable for anaphoric and ostensive use, and for use as a personal pronoun. Its most frequent use, as attested by corpus data, is as a personal pronoun, notably in the genitive (rendering the possessives “his, her, their”). Lyons (1999, p. 134) calls personal pronouns the “pronominal counterpart of definite articles”, and adds that just like articles, personal pronouns typically derive from demonstratives, something OCS would then be a clear example of with both *онъ* and, possibly, **ѣ*.

For the full declension of the pronoun/demonstrative **ѣ*, see table 1 on the next page.

⁹ Translation from: http://clover.slavic.pitt.edu:8080/exist/paul/data/paul_main.html (25-04-2013)

¹⁰ Another demonstrative, *онъ* (that there) is classified by Flier as “[representing] a different perspective, relating more to mutual opposition than to personal proximity” (Flier, 1974, p. 59). *Онъ* is found in uses such as “some say..., others...” and “they struck *one* and killed *another*”. In the PROIEL corpus there is no example of this demonstrative used attributively.

Table 1 Full declension of the pronoun **jb*

| Sg. | M | N | F |
|------------|----------|----------|----------|
| N | -*jb | -*je | -*ja |
| A | -jb | -je | -jɔ |
| G | -jego | ← | -jeje |
| D | -jemu | ← | -jeji |
| I | -jimь | ← | -jejɔ |
| L | -jemь | ← | -jeji |
| Du. | M | N | F |
| NA | -ja | -ji | ← |
| GL | → | -jejo | ← |
| DI | → | -jima | ← |
| Pl. | M | N | F |
| N | -*ji | -*ja | *-je |
| A | -je | -je | -je |
| G | → | -jichь | ← |
| D | → | -jimь | ← |
| I | → | -jimi | ← |
| L | → | -jichь | ← |

Based on the un-transliterated chart in Čaburgaev (1974, p. 215).

To form LF adjectives/participles, the forms in table 1 are added to the basic form of the adjective/participle, which itself declines like a noun. These combined forms then undergo various reductions and assimilations, and by the time of OCS (i.e. the Gospel texts), the pronoun has merged completely with the word, becoming a flexional ending. The following examples (masculine singular) from Čaburgaev (1974, p. 230) illustrate this. Notice how the SF base also declines.

- | | |
|---|---------------------|
| (9) N dobrъjb > dobryj > dobry | (short form: dobrъ) |
| G dobrajego > dobraego > dobraago > dobrago | (short form: dobra) |
| D dobrujemu > dobruemu > dobruumu > dobrumu | (short form: dobru) |
| L dobrějemь > dobrěemь > dobrěmь | (short form: dobrě) |

Table 2¹¹ on the following page illustrates the compound forms of SF adjectives + *jb*. There are some further variations on these endings. The columns show both short and long form.

¹¹ This is a modified version of the table in Flier (1974, p. 66), combined with the soft/hard long form table in Čaburgaev (1974, pp. 232-233).

Table 2 Short form and long form endings of adjectives and participles, soft/hard, M|N

| | Case | SF – M N | SF – F | LF – M N | LF – F |
|------------|------|----------|--------|---------------------|-------------|
| Sg. | N | ъ/ь о | а | ъѣ>ѣ(i),ii оје/еје | аја, ајѣ |
| | A | ъ/ь о | ѡ | ъѣ>ѣ(i) оје/еје | ѡјѡ |
| | G | А | ѣ | ајѣго | ѣје/ѣје |
| | D | У | ѣ | ујѣму | ѣји/ји |
| | I | омъ/емъ | ѡѡ | ѣмъ/iimъ | ѡјѡ,ѡјѡ/ѣјѡ |
| | L | ѣ | ѣ | ѣјѣмъ/ijimъ | ѣји/ји |
| Du. | NA | а ѣ | ѣ | ајѣ/аја ѣји/ји | ѣји |
| | GL | У | у | ујѣ/ујѡ/ѡјѡ | ујѣ |
| | DI | ома | ума | ѣма/ijima | ѣма |
| Pl. | N | i а | ѣ | ји ајѣ | ѣје/ѣје |
| | A | ѣ а | ѣ | ѣје/ѣје ајѣ | ѣје/ѣје |
| | G | ъ | ъ | ѣјѣхъ/ijichъ | ѣхъ |
| | D | омъ/емъ | амъ | ѣмъ/ѣјѣмъ/ijimъ | ѣмъ |
| | I | ѣ | ами | ѣми/ѣјѣми/ими/ijimi | ѣми |
| | L | ѣхъ | ахъ | ѣхъ/ichъ | ѣхъ |

Certain difficulties present themselves in the analysis of these forms. For example, the difference between SF and LF in the instrumental feminine singular can be nonexistent according to table 2. As is seen in table 1, the pronominal form is *jejѡ* and not the hard *-ojѡ*, but this will depend on the stem of the adjective in question. Rondestvedt (1986, pp. 86-87) lists the LF as *-ѡјѡ*, and the short form as *-ojѡ*, whereas Ivanova (1977, pp. 119-120) lists both *-ojѡ*, *-ѡјѡ* and *-ejѡ* as LF. The feminine is not the most frequent gender in the biblical texts, so this is not something that affects the data greatly¹². The dative plural also shows small differences between SF and LF, as the SF already has the *-(*)mъ* ending, and for certain soft adjectives SF and LF are practically impossible to tell apart, such as for *božii*, “of-God/divine”, *velii*, “great”, *divii*, “wild”.

In transliteration, the pronoun **jъ* can be rendered in different ways. The Glagolitic alphabet¹³ used three different letters interchangeably for the sound /i/, and Cyrillic transliterators made use of four variants: и, i and occasionally ѣ or ѣ (Lunt, 2001, p. 23). The last two are never found in original Cyrillic texts, and in normalized texts, и and i are used (Ibid., p. 24). These various letters are then transliterated into latin script in three different ways: *jъ*, *ji* and *i*, depending on phonetic surroundings and some rules pertaining to function.

In the literature on OCS grammar, **jъ* attached to an adjective or participle is referred to differently by different authors. Leskien (1992, p. 119) refers to it as a “postposed article”, Flier (1974, p. 67)

¹² In the entire corpus, 10.8% of the adjectives/participles are in the feminine and 0.4% are feminine instrumental singular.

¹³ Codex Marianus, the main source for material used in this thesis, was originally written in Glagolitic and then transliterated into Cyrillic.

does not want to call the LF and SF definite and indefinite respectively, Lunt (2001, p. 65) refers to LF as a “compound or adjectival declension” which is “roughly equivalent to the English definite article”. Čaburgaev (1974, p. 229) also chooses to describe the compound form as fulfilling the function of an article that adds nuances of meaning to the adjective:

“...указательное местоимение, присоединенное к именной форме прилагательного, первоначально выполняло функцию определенного члена (артикля), вносящего дополнительные оттенки в значение прилагательного”.

Čaburgaev (1974, p. 229) specifies that the LF was used to individualize the noun in question, and that the speaker would use LF for adjectives when ascribing to the noun a quality that he considered to be already known to the addressee. However, if the head noun was in no need of any specification (i.e. if it was so well known that it was considered superfluous to further mark it) or if the information given about the head noun was simply new, SF was used¹⁴. Whether or not this is true is something that will be examined in chapter five. It should also be noted that Borodič (1963, p. 162) dates the provenance of the LF to pre-Slavic times, saying that it initially had the meaning of definiteness, but that in the “historical period” (i.e. the period we have manuscripts from) the SF/LF distinction started disappearing. This is further discussed in chapter 3.6.

3.2. **тъ** - definite article or demonstrative?

In an article from 2009, Dimitrova-Vulchanova and Vulchanov present a view that differs from traditional OCS linguistics. The OB¹⁵ texts they use are different from those used in this thesis, with the exclusion of Codex Suprasliensis and the original Greek New Testament¹⁶. In the article, the authors present **тъ** and **сь** not solely as demonstrative pronouns, but also as definite articles in OB. In their account, word order is of key importance for definiteness. They claim that the pre-nominal position is primarily used for modifiers in *indefinite* expressions (Dimitrova-Vulchanova & Vulchanov, pp. 80,84), whereas the post-nominal position shows a strong tendency to incorporate phrasal modifiers in *definite* nominal expressions (Ibid., p. 78). They refer to LF adjectives as “expanded forms of adjectives”, that can be expected to be found in the post-nominal position (Ibid., p. 78), but they never equal it to a definite marker¹⁷. In their account, **тъ** as a definite article is a phrase-second clitic, meaning it will always appear in second position within the definite NP (Ibid., p. 81). If an adjective precedes the head noun in a NP, the structure Adj > Art > N is achieved. Conversely, N > Art

¹⁴ «...когда указание на признак (свойство) предмета не требовало подчеркивания его известности, специфичности или просто было новым в сообщении» (Čaburgaev, 1974, p. 229).

¹⁵ The authors use the term Old Bulgarian for Old Church Slavonic.

¹⁶ Texts they use that are different are: The Enin Apostle, the Symeon Miscellany, the Dobromir Evangelium, Mark, the Psalterium Sinaiticum. These are later texts than the Gospel texts.

¹⁷ Rondestvedt (1986, p. 110) says that “The history of adjectives in Bulgarian shows a semantic confusion of the two forms [SF & LF] leading toward the development of a new article from the demonstrative **тъ**.”

> Adj is equally possible. This can be compared to the definite article that has evolved in Bulgarian, which attaches either to the noun or the leftmost adjective, as seen in example (11). (10) is an example from Codex Suprasliensis from the 2009 article (Ibid., p. 81), showing the same structure:

- | | | |
|------|---|---|
| (10) | blaženyjъ tъ glasъ tēs makarias fōnēs tautēs | (CS 123, 28-9) |
| (11) | Kniga- ta Interesna- ta kniga | “the book” “the interesting book” (Dost & Gribanova, 2006) |

Tъ in example (10) appears in second position and follows the LF adjective *blaženyjъ*, which, since it appears in pre-nominal position, does not conform to the general structure of indefinite modifiers in pre-nominal position (as presented by the authors of the article). Dimitrova-Vulchanova and Vulchanov identify *tъ* with the Greek definite article *tēs* (f.gen.sg.) and not with the Greek demonstrative *tautēs* (f.gen.sg). This differs from the traditional view accepted among grammarians, according to which the Greek definite article corresponds to the **jъ* addition on the LF adjective (*blaženy-jъ*), and *tъ* corresponds to the demonstrative *tautēs*, as represented below (underlined and bolded words correspond to each other).

- | | | |
|------|---|----------------|
| (12) | blaženyjъ <u>tъ</u> glasъ tēs makarias fōnēs <u>tautēs</u> | (CS 123, 28-9) |
|------|---|----------------|

This is a comparison that, among others, Archaimbault (1992, p. 214) presents, saying that “this one [the enclitic pronominal element **jъ*] has been compared to the article that accompanies the adjective in Greek”¹⁸. Vaillant (1942, p. 1) categorically begins his article “L’article en vieux slave” with the words “the OCS article is not the post-positioned *tъ*, which in OCS remains a pure demonstrative”¹⁹. Kurz (1963), concluded, after studying some 1541 OCS examples of postposed *sb*, *tъ* and *onъ* that these are as a matter of fact demonstratives, not articles. Dimitrova-Vulchanova and Vulchanov therefore present opinions that are quite unusual, at least in the Western tradition, and they provide no statistics or numbers for their claims.

A possibly interesting parallel is to look at this system in the light of the Scandinavian languages, in which the use of a similar structure to that OCS displays is found in various forms. In Norwegian, and partly in Swedish, definite NPs together with a definite adjective must be accompanied by a definite article/demonstrative, rendering a system of threefold marking of definiteness. The definite article does not differ from the demonstrative other than in being prosodically less prominent (Dahl, 2004, p. 159).

¹⁸ My translation of “Celui-ci a été comparé à l’article qui accompagne l’adjectif en grec.”

¹⁹ My translation of “L’article vieux-slave n’est pas *tъ* postposé, qui conserve en vieux slave sa valeur pleine de démonstratif.”

| | | |
|------|---|--|
| (13) | Den röda boken (The red book) | (art-DEF + adj-DEF + noun-DEF) |
| | <i>Den röda boken</i> (<i>That</i> red book) | (<i>den</i> as a demonstrative, pointing) |
| | Den där/här röda boken | (This/that red book) |

This is referred to as “double determination” (Lyons, 1999, p. 78). All markers need not be present at the same time in for example Swedish; the important part is that *some* marker of definiteness is present (Språknämnden, 2005, p. 114), and the *adjectival* marking seems to be the central point in this system. It would therefore not be unheard of that *t̃b* could be viewed as both a definite article and as a demonstrative in OCS, just like *den/det* in the Scandinavian languages.

Furthermore, Dimitrova-Vulchanova and Vulchanov (2009, p. 82) state that “in NT Greek the article can, and does cooccur with the demonstrative in a variety of construction types, while such cases are not attested in our OB data”. They take *t̃b* and *s̃b* to be demonstratives (and not articles) when they occur in the pre-nominal position or when they appear post-nominally after the head *and* its modifiers. This is also illustrated with an example (Ibid., p. 82):

| | | | |
|------|-----------------|------------------------------|---------------|
| (14) | pre-nominally: | въ tyję dni | (DE, 14b, 8) |
| | post-nominally: | pri narodě ṽsemъ sem | (CS. 149, 15) |

These ideas make it interesting to also investigate the role word order plays for the appearance of SF and LF, and the placement of the definite article/demonstrative. Does the PROIEL data support the idea of indefinite pre-nominal expressions? If OCS word order is found to just replicate Greek word order in our texts, as Flier (1974, p. 36) indicates, this may not be of very much interest. This is investigated in chapter 5.1.8. The general take on *t̃b* and *s̃b* here, though, is that they are demonstratives and not articles. They are treated in cooccurrence with adjectives and participles in chapter 4.3.3.9, where it will become obvious that *t̃b* very rarely co-occurs with adjectival modifiers. Instead, what we find is *s̃b*. If *t̃b* was truly a definite article, one would expect to see it more often, unless the claim is that it only appears together with NPs that have no other modifiers, but the authors make no such claim. They use later texts, which may explain the differences. *T̃b* is, however, mentioned by Vlasto (1986, p. 133) as frequently appearing post-posed on nouns in OCS, and as being the source of the Bulgarian article innovation. He also mentions that it is more frequent in this position than *s̃b*. What is stated here regarding the behavior of *t̃b* is only that it rarely cooccurs with adjectival modifiers.

3.3. The work of Michael Flier

Possibly the most extensive work on short/long form adjectives and participles is that of Michael Flier (1974), the main points of which will be resumed here.

In his work, Flier divides adjectives (and participles) into different categories, specifying what general rules apply in each category, and then proceeds to analyze instances where the rules are seemingly broken. In cases where unexpected forms appear, he often attributes this to Greek influence or explains them as being stylistic variations. As for the provenance of the LF marking, Flier (1974, p. 68), inspired by Vaillant, argues that the development of LF in (16) was motivated by (15):

- | | | |
|------|------------------------|-------------------|
| (15) | člověkb iže estb dobrb | “man who is good” |
| (16) | dobry člověkb | “the-good man” |

In (15), the pronoun **jb* is proclitic in the relative pronoun *iže* (*jb* + particle *že*), whereas in (16), it has become attached as an enclitic at the end of the adjective, rendering *-y* (*-jb*)²⁰.

The concatenation of regular adjectives and *j-* was no doubt motivated by the close semantic tie existing between attributive adjective and relative clauses with predicate adjectives (Flier, 1974, p. 68).

Flier points out that, even though there is a “semantic tie” between the types of sentences in (15) and (16), the two are *not* identical, as Vaillant (1942, p. 5) states. The relative clause does not mark the head noun as definite or specific, it merely delimits it (Flier, 1974, pp. 69, 80). The LF adjective, on the other hand, renders the noun definite or generalized (only in the singular, according to Flier) but not specific. He argues that the LF cannot cooccur with *něky* or *eterb*, which are markers of specificity, and thereby LF cannot cooccur with indefinite specific nouns (Flier, 1974, p. 69). This does, however, happen in younger texts, such as the Codex Suprasliensis, but there is no support for this in the small amount of texts from CS currently present in the PROIEL corpus.

Flier (1974, p. 67) does not consider **jb* to be a definite article since it can only mark definiteness for regular adjectives²¹, as well as for certain adjectives with the *-bšk* suffix. As we shall see further on, not all adjectives seem to form LF, and the grammatical definite marker is therefore limited in scope. Flier thereby has a criterion that the marker of definiteness must be applicable to *all* adjectives, regardless of type, in order to qualify as an “article”.

The division of adjectives that Flier makes is the following:

- Adjectives of affiliation and apposition
- Adjectives derived from or modifying nouns with unique reference
- Adjectives modifying common nouns or undergoing nominalization
- Nominalizations

²⁰ When followed by *-j-*, */b/* is in a “tense” position and varies freely with */y/*. See Lunt (2001, p. 35).

²¹ Regular adjectives are adjectives that can occur in both short and long form. This term will be used throughout the paper.

Based on Flier's research, OCS nouns can be divided into those that show SF or LF due to their inherent semantic definiteness, and those that vary depending on the context.

Flier expects SF to appear for most of what he refers to as "adjectives of affiliation and apposition" (i.e. adjectives formed from particular suffixes, further discussed in chapters 3.5.1 and 5.1.1). Flier (1974, p. 81) explains this with the fact that such constructions are not derivatives of relative clauses, and therefore no LF has been developed for these adjectives. A more plausible argument that he also presents is that the semantic nature of the nouns underlying such adjectives explains their use of SF. These nouns are often proper nouns, which in themselves are individualized, making any morphological marking of definiteness "redundant" (Ibid., p. 100). This does not mean that they never show LF, and whether such marking is actually "redundant" is debatable (see chapter 2.2.1 on economy/explicitness). According to Flier, attributive adjectives of affiliation in -ьск, in particular, show vacillation when derived from proper nouns (which are unique), but are limited to the SF when derived from common nouns (which are non-unique). When derived from nouns referring to unique entities, -ьск adjectives are *mainly* long (Ibid., p. 137), and, in general, LF is expected from what Flier calls "adjectives derived from or modifying nouns with unique reference"²². Adjectives in this category are referred to by Flier (Ibid., p. 103) as "U-adjectives". Examples he uses are *istinъn-/istovъ* (the true X), *inočędъ* (referring to Christ, "the-only-begotten"), i.e. adjectives that modify unique nouns, and adjectives derived from unique nouns such as *nebesъskъ* (of-Heaven). Adjectives such as the latter are, of course, also adjectives of affiliation, meaning that these two categories overlap. These categories easily become somewhat confusing. Flier (Ibid., p. 127) also places ordinal numbers and "polar opposites" in this U-category, since they denote a unique part of a whole: the first, second, third, etc.; left vs. right. Flier (Ibid., p. 103) attributes SF among these adjectives to phonetic or graphemic interference or external Greek influence.

As for regular adjectives that are neither relational nor modifying unique entities, SF/LF is expected to depend on the pragmatic features of the nouns they modify, as "common nouns are not taken to be unique in any way prior to their inclusion in a particular discourse and their referents are therefore not known ahead of time by the speakers of the language" (Flier, 1974, p. 137). Common nouns are divided in the following manner by Flier:

- As members of a general set of similar nouns and as therefore representing the referential feature [INDEF]
- As particular members of a set of similar nouns, able to expand the underlying hierarchy to [SPEC] and to allow for the cooccurrence of *něky* or *eterъ* ("some", "a certain")
- As individualized members of a set of similar nouns, able to expand the underlying referential hierarchy to [DEF] in cooccurrence with regular adjectives

²² Combining these two groups is not necessarily a very good idea, especially not for the sake of clarity.

- As generalized nouns, each one able to represent the entire set of nouns of which it is a member, expanding its underlying hierarchy to the feature [AGGREGATE] in cooccurrence with regular adjectives

According to Flier, LF should be found on the adjectives and participles modifying *definite* nouns, which are further divided into vocatives, anaphoric nouns with or without immediate antecedent and nouns occurring with deixis. Flier (1974, p. 138) further argues that, since LF is also found with nouns that do *not* have the underlying feature [+DEF], notably with generalized nouns in the singular, as well as with plural nominalizations referring to indefinites, LF cannot be referred to as “definite”. This is a strange claim; that generic nouns should not be considered definite is questionable (see chapter 2.2.3). Definite articles are used in many languages to denote generic referents. Here it is also pertinent to remember the distinction between semantic and grammatical definiteness discussed in regards to Lyons, as these two categories may or may not overlap.

OCS nominalizations are often treated separately from attributive adjectives since they are suspected of showing more LF than the context would have them do. For nominalizations, Flier sees a definite trend towards LF plural, regardless of whether the nominalization is actually definite or not. This has then had an effect on plurals in general, giving rise to undue vacillation (Flier, 1974, p. 164). Especially the latter is a claim that will be examined in the analysis chapter of this thesis. Flier’s claim that LF only renders nouns generalized in the singular will also be examined.

3.4. The work of Karen Anne Rondestvedt

Rondestvedt, in a PhD thesis from 1986 entitled *Definite/indefinite and related pragmatic categories in Early Original Slavic*, examines the role of SF/LF in a group of Early Original Slavic²³ texts, all written originally in Slavic with no translations from Greek. The texts she uses include all of the longer Early Original Slavic texts²⁴ that the author had knowledge of, stemming from Bohemia, Moravia, Pannonia and Bulgaria. The criteria for choosing these texts were that their originals date from before the fall of the Bulgarian Empire in 1018, and that the oldest extant manuscript be dated to the 15th century or earlier (Rondestvedt, 1986, p. 8). That these texts have not been used much in previous research is explained by the author by the lateness of the extant manuscripts, which have obviously been contaminated by the language of their scribes (Ibid. p. 5). The styles in these texts vary from “very literary, often with much imitation of Greek, to rather informal”. The content goes from saints’ lives to law code and includes both prose and poetry (Ibid. p.8). All in all, her material constitutes

²³ This is Rondestvedt’s own term for these texts (1986, p. 5).

²⁴ The texts used are the following: the Life of Methodius, the Life of Constantin, Poxvala to Cyril and Methodius, Anonymous sermon in the manuscript Clozianus, the Life of Wenceslas, 6 sermons by Clement of Oxrid, “O pismeněxъ” by the monk Xrabr, Presbyter Kozma’s “Treatise against the Bogomils”, 3 miracles from “Tale of the Iron Cross” and Incription from the tomb of Mostič (Rondestvedt, 1986, p. 10).

approximately 160 pages. The mixed nature of her texts could be seen as a problem in finding patterns, but considering the size of the EOS text corpus, such a mixture was obviously deemed necessary. The purpose in choosing exclusively EOS texts was (partially) to study word order more thoroughly, which is difficult with translations from Greek that rarely deviate from the original. Furthermore, Rondestvedt's method consists in going from meaning to form – and *not* the other way around – as she does not look uniquely at, for example, SF/LF as indefinite/definite, but looks for *a* system of definiteness marking, regardless of what it may end up consisting in.

In her thesis, Rondestvedt (1986, p. 226) shows that the use of SF/LF in Early Original Slavic is essentially the same as that previously observed in OCS. There is no single factor, such as definite/indefinite, that can explain *all* appearances of SF/LF in EOS (Ibid., p.111), and the author claims that the same is true for OCS.

Rondestvedt's approach is rather similar to that of this thesis, notably focusing on definite/indefinite as a pragmatic category. She does not, like Flier, look at the idea of definiteness on the word or sentence level, but in the greater context of an entire text with a left and right context. As such, her work contains much less comments on the semantics underlying various groups of words, which is what Flier often relies on. Furthermore, even though she mentions Flier's work repeatedly, Rondestvedt (1986, p. 111) is not interested in comparing EOS to OCS, but in examining whether definite/indefinite is systematically expressed somehow in the EOS texts, be it with the use of SF/LF, word order or a combination of both. She is not persuaded by Flier's statement that the plural has neutralized the SF/LF distinction (supposedly influenced by the behavior of nominalizations, which according to him show more LF than expected), and this is a claim that will be questioned in the examination of the PROIEL data as well.

Another point worthy of mention is that Rondestvedt had the intention of checking the distribution of adjective forms for indefinite, definite and generic referents (in correlation with sentence position and topic/comment) statistically, but was unable to do so due to uncertainty regarding the status of the referents. There were too many occurrences where she found it impossible to tell the status of a referent, and therefore she decided to drop statistical data, and her discussion is therefore "impressionistic and not quantified". She does count instances for each category she examines from randomly chosen texts in order to give suggestive numbers (Rondestvedt, 1986, pp. 112-113).

In the EOS texts, what interferes the most with the SF/LF as an expression of indefinite/definite is morphology, i.e. suffixed adjectives, and word order plays very little role as a system to mark definiteness (Rondestvedt, 1986, p. 228). The findings in chapter five will show that the OCS data display similar results.

3.5. Remarks on adjectives

3.5.1. Types of adjectives and suffixes

There are different types of adjectives. The discussion on *what* exactly adjectives are (i.e. what separates them from verbs and nouns) belongs elsewhere, but some basic distinctions are useful. The nature of adjectives varies from language to language, so statements made here are only valid for the languages discussed (Slavic, Germanic).

First of all, we can separate *qualitative adjectives* that ascribe a quality to the noun it modifies (*red*, *big*) from *associative adjectives* that associate the noun with another noun by the means of turning the second noun into an adjective. *Possessive adjectives* are a subset of associative adjectives, an example of which would be the Russian *sobačij ošejnik*, “dog collar”. Some, like Tolstoy (1957), separate the category into *possessive* and *relational*²⁵, where possessive adjectives (*Isusovъ*, “of Jesus”) are expected to show SF and relational adjectives generally occur in the LF. Rondestvedt (1986, p. 95) further explains this division: relational adjectives that are expected to display LF are adjectives that carry a temporal (*utrěi*, “morning”), locative (*desnyi*, “right”), part-whole (*сръкѣньни*, “church-”) meaning, and that indicate an intended function (*купѣньни*, “trade-”). Relational adjectives relate the noun to some thing or other, but do not thereby render the noun automatically definite. Possessive adjectives, on the other hand, do (cf. DG systems discussed in chapter 2.2.2). The qualitative adjectives are the ones that, in the words of Rondestvedt (ibid., p. 95), “are the least able to individuate their head nouns by their lexical meaning”, meaning that it is for this kind of adjective that we would expect to find the most variation in SF/LF, whereas the opposite is true for associative adjectives (in particular possessive ones), which “anchor” the referent to another referent.

The relations expressed by associative adjectives are, naturally, rendered in different ways according to language: Ancient Greek uses genitive constructions, English and Scandinavian languages can use both genitive constructions and nouns as attributive modifiers on other nouns (compound words), and OCS frequently makes use of possessive adjectives. The following example from (Matt 30:26) is used by Michael Flier, here supplemented by Swedish and English translations:

(17) Matt 30:26:

1 вѣспѣвъше изидѡ въ горѡ **елеонѣскѡ**
kai humnēsantes exēlthon eis to oros **tōn elaiōn**

And when they had sung a hymn, they went out
to the Mount **of Olives**.

Adjective formed from eleonъ with the -ъsk suffix
Genitive plural of ho elaios (wild olive) or hē elaiā
(olive tree)

Genitive construction.

²⁵ The Russian split of *prizhāatel'nye* and *otnositel'nye prilagatelnye*.

När de hade sjungit lovsången gick de ut till
Olivberget²⁶.

Compound word.

OCS also makes use of genitive constructions, as in *gybel' chriz'mnaja* vs. *gybel' chrizmy*, “the loss of nard” (Le Guillou, 1984, p. 198). The point of interest here, however, is that this type of possessive and relational adjectives are formed by the use of a range of suffixes in OCS, making them rather easily identifiable. These are suffixes that have been observed to appear on SF possessive adjectives, regardless of context, the explanation for which would be that the underlying semantics sufficiently individualize the noun, rendering an additional LF distinction superfluous. Čaburgaev (1974, p. 229), among several others, states that possessive adjectives are *almost* never used in the LF. According to Borodič (1963, pp. 183-186), the following distribution²⁷ of the suffixes²⁸ can be established:

| | |
|--------------|--|
| -ov | Spread from gen.pl. of ŭ-stems and became a marker of possession when formed from animates. |
| -in | Formed from proper nouns and generally known individuals, as well as from plants and animals. |
| -jb | Same as -in. |
| -b'sk | Somewhere in between possessive and qualitative adjectives. Started out as individual possession, replacing genitive noun constructions. |

Zverkovskaja (1986, p. 88) qualifies these suffixes, with the exception of -b'sk, as having more or less the same function as possessive genitive constructions. Brodowskaja is of the opinion that -b'sk adjectives *are* the equivalent of such constructions when short (not when long), but there is not widespread agreement on this matter (Borodič, 1963, p. 187). As becomes clear from this enumeration, these adjectives are predominantly formed from animates, something that is also confirmed by Flier (1974, p. 82), and motivates the SF on the basis that LF is *redundant* (see chapter 2.2.1):

The relationship between head noun and animate nouns is tighter than that between head nouns and inanimate nouns since in many cases true possession may be involved among the former, probably the strongest form of affiliation.

Borodič (1963, p. 192) claims that adding an -ov or -in adjective to a noun *always* makes the noun definite, which points towards these adjectives functioning like DG-adjectives.

²⁶ <http://www.bibeln.se/las/2k/matt#q=Matt+26%3A30> (2012-03-22)

²⁷ Le Guillou (1984, p. 100) presents another distribution, based on the phonetic nature of the underlying noun that the adjective has been formed from. As such, -in and -jb differ in that the first is formed from nouns ending in -a and -b, whereas the second is formed from nouns ending in -b.

²⁸ These suffixes are rendered somewhat differently in different sources due to the difficulties related to the transcription of the Cyrillic letter «и» : -ov, -in, -i (jb), -b'j and -b'sk in Borodič (1963, p. 183), -ov, -in, -j, -ij in Flier (1974, p. 82).

In the PROIEL corpus, non-qualitative adjectives are annotated for a slightly more detailed array of suffixes: *-ij*, *-in*, *-j*, *-n*, *-nj*, *-ov* as well as *-sk*. The *-sk* (*-ḅsk*) suffix is the one that stands out the most, since it shows the greatest fluctuation between LF and SF, and for this reason it is looked closer at in chapter 5.1.1. Qualitative adjectives can also show these suffixes, but they do not create a relational/possessive tie between head word and the noun the adjective is derived from, so the head word is not automatically individualized. For this reason, these adjectives have not been tagged for suffixes in the corpus.

Flier (1974, p. 82) argues that where these suffixed adjectives appear in the LF, it is a result of the change towards LF attributive adjectives. Table 3 shows the complete amount of suffixed adjectives in the corpus (the total amount of adjectives is 3819), showing that some of these suffixes are not very frequent. The part of the corpus tagged for IS has not been singled out in this sample, which will be the case in the analysis chapter.

Table 3 Suffixes and their distribution in the PROIEL corpus

| Suffix | Total | Percentage ²⁹ | SF | LF |
|--------|-------|--------------------------|-------------|-------------|
| -ov | 278 | 7.3% | 278 (100%) | 0 (0.0%) |
| -in | 31 | 0.87% | 29 (93.5%) | 2 (6.5%) |
| -ij | 163 | 4.7% | 161 (98.8%) | 2 (1.2%) |
| -j | 69 | 1.85% | 69 (100%) | 0 (0.0%) |
| -n | 261 | 6.8% | 91 (34.9%) | 170 (65.1%) |
| -nj | 42 | 1.13% | 33 (78.6%) | 9 (21.4%) |
| -sk | 381 | 10.1% | 148 (38.8%) | 233 (61.2%) |
| | 1225 | 32.8% | 809 (66.0%) | 416 (34.0%) |

This table shows the total number of instances for each suffix, the percentage this makes up out of all of the adjectives in the corpus, and the distribution of SF/LF among said suffixes.

As becomes obvious from table 3, certain of these suffixes can be excluded straight away. To narrow down the data, the suffixes *ov*, *j*, *in* and *ij* have been removed. In the narrowed down final dataset, the total percentage of LF is 58.2%, meaning that the *-ḅsk* and *-n* suffixes are very similar in their distribution of LF and SF to the entire corpus of regular adjectives. The *-n* suffix shows a distribution of SF/LF in relation to IS-status that closely mirrors that of regular adjectives, i.e. LF where we expect it (OLD, ACC, etc.), and therefore it will not be investigated more closely. The *-ḅsk* suffix, on the other hand, does not show a similarly clear-cut distribution. It is examined separately in chapter 5.1.1.

²⁹ This is the percentage of all adjectives in the corpus.

3.5.2. Comparatives and superlatives

As was pointed out in the chapter on definiteness, superlatives in English and other languages that mark definiteness are always accompanied by definite articles (alternatively possessive pronouns). Dahl (2004, p. 153) says that “the noun phrases they are used in normally have definite reference by virtue of their meaning”. In a series of something, there is usually only one referent or group of referents that is “the worst”, “the slowest”, etc., so these can be singled out from the larger, indefinite mass.

The mention of degrees for adjectives is of some relevance for the distinction between SF and LF. In OCS, the comparative is formed by adding **-jbs* (**-jes for sg.nom.neut*) to the root of the positive adjective, either directly or by means of the vowel *ě* (Čaburgaev, 1974, pp. 231,234). As a result, some of the bare comparative forms (i.e. SF) are indistinguishable³⁰ from LF comparative, notably the ones ending in *-ěi* and not *-ii* in the nom.sg.masc/neut (Ibid., pp. 236-237). Compare the nom. and gen. endings for masc/neut sg. in example (18):

| | | | |
|---------------------------|----------------|----------------------|----------------|
| (18) SF m/n. n. sg. comp: | starěi/starěje | LF m/n .n. sg. comp: | starěi/starěje |
| SF m/n. g. sg. comp: | starěišja | LF m/n. g. sg. comp: | starěišjajego |

Flier (1974, p. 148) suggests that the distinction between SF and LF was neutralized in the comparative, at least in some texts, whereas Rondestvedt (1986, p. 132) considers them to follow the expected pattern for definite-indefinite-generic. As for the superlative, OCS mostly made use of the comparative to fill this function as well. Rondestvedt (Ibid., pp. 116-117) says that Early Slavic had no formal superlative and that LF comparatives can be expected to be actual superlatives, but that this is hard to determine from the material. Čaburgaev (1974, p. 238) presents some alternative variants of superlatives, with the modifiers *вѣсь*, *вельми* as well as the rare forms in *nai-* and *pre-* + comparative, but says that their use is not very frequent. Since the translators who worked with Greek texts must have known what the Greek superlative was, it is easy to check (using PROIEL data) how many of the OCS comparatives correspond to Greek superlatives and whether or not there is a difference in use of SF and LF.

The PROIEL corpus only contains 17 LF attributive adjective comparatives that have Greek originals. Among these, 15 correspond to similar adjectival constructions in Greek (the remaining two use adverbs), of which 73.3% are Greek superlatives. Among the nominalizations, the LF comparative/Greek superlative overlap is 9/11 (81.8%). Rondestvedt's theory that OCS LF

³⁰ When the form of an adjective is not indicative of whether it is SF or LF it is tagged as uncertain in the PROIEL corpus. Such instances are not included in the data used in chapter 5. 18.8% (13) of comparatives fall into this group.

comparatives act as superlatives is therefore supported by the PROIEL data. The number of examples is rather small, but the combined results for attributes and nominalizations are statistically significant³¹.

3.6. Changes in OCS grammar in the 10th and 11th centuries regarding SF/LF

The development that some believe eliminated the distinction between LF and SF is supposed to have begun in the 10th or 11th century. It can be said to have culminated in the situation found in modern Russian, where SF adjectives are limited to the nominative and are found predicatively. Attributively, and in other cases, regular adjectives display LF. Flier (1974, p. 8) refers to this development as “pressure from native Slavic trends towards the establishment of long-form attributes and short-form predicates”. Something similar is found in Borodič (1963, p. 163), who argues that the system of marking definiteness using SF/LF dissolved under the influence of “new, specific Slavic categories” such as animate/inanimate, aspect and the difference between attributive and predicative position. In the 10/11th century manuscripts in the PROIEL corpus (of which the majority are copies of earlier 9th century manuscripts), qualifying predicative adjectives and participles are short, and only the attributive, nominalized and appositive ones show SF/LF variation. There are also instances of LF in identifying predicative constructions, but they are not very numerous. The SF for predicatives in OCS is explained by them being non-referential; as they do not refer to anything that is identifiable or known there is no need for them to display LF. What is found in the predicative position is the comment, which often presents new information about an entity that is either already known or which is being introduced. Since the information is often new (qualifying), the unmarked SF is the more logical choice. There is no semantic or pragmatic reason why they should ever have displayed LF. Therefore, the “Slavic trend” should only touch the attributive adjectives and participles.

The system of definite LF has survived in Serbo-Croat, but is disappearing, and Macedonian and Bulgarian have instead developed a clitic definite article (Lyons, 1999, pp. 82-83). Furthermore, Tolstoy (1957, p. 50) remarks that in Serbo-Croat, the LF is taking over an increasing amount of lexical groups in *both* attributive and predicative positions. He also argues that since the change towards the attributive/predicative split is only supported by the development in Russian it cannot be the “cause” for the dissolution of the definite/indefinite categories. The other East-Slavic languages, Ukrainian and Belorussian, do not entirely show the same clear pattern as Russian. Ukrainian has altogether done away with LF adjectives for N/F singular nominative and accusative, as well as M/N/F

³¹ P-value = 0.009355, using a binomial test.

plural nominative, and has a group of fully short adjectives similar to the Russian ones. Belorussian has separate SF and LF for certain adjectives, otherwise only long. Rondestvedt (1986, p. 110) is also opposed to the idea of the predicative/attributive split as an explanation for this development since “the trend toward long-form attributive and short-form predicative adjectives did not and does not exist in South Slavic”. She argues that this development was later and limited to East and West Slavic, and cannot be used to explain anything in OCS.

Since CS is a younger text, we could expect any supposed development to be more apparent, with an increasing amount of LF attributes that are not necessarily definite. On the other hand, presently, the IS-tagged texts from CS only make up a small part of the corpus. The majority of the texts come from Codex Marianus. This makes it difficult to say anything conclusive regarding this trend.

3.7. Concluding remarks

This chapter has presented some essential background information necessary to the understanding of the problem of SF/LF in OCS. Certain questions that will be addressed in the analysis in chapter five have arisen throughout the chapter. More specifically, these are the following:

- Are highly familiar referents marked by SF adjectives because using LF would be superfluous, as Čaburgaev indicates?
- Is word order significant for definiteness? Is the pre-nominal position primarily occupied by indefinite modifiers and the post-nominal by definite modifiers, as Dimitrova-Vulchanova and Vulchanov claim?
- Is it true that LF only renders singulars generalized, not plurals, as Flier claims?
- Is LF on adjectives derived from proper nouns so redundant that it almost never appears, as Flier seems to indicate? And do -ьsk adjectives derived from proper nouns display vacillation whereas they are always short when derived from common nouns, as Flier says, or is there a pragmatic reason for their use of SF/LF?

4. Method

In this chapter, the sources from the PROIEL corpus will first be presented in section 4.1, after which the IS annotation scheme used on them is explained in section 4.2. The way the data has been handled as well as all exclusions that have been carried out prior to the final analysis in chapter five can be found in section 4.3.

4.1. Sources

The language referred to as OCS is usually considered the language found in non-East Slavic manuscripts dated prior to the year 1100. This is, according to Lunt (2001, p. 4), a “convenient (but

arbitrary) date". None of the earliest mid-9th century biblical manuscripts, for which OCS was originally created, have been preserved to this day. Instead, the Gospel texts present in the PROIEL corpus are 10th/11th century copies of these manuscripts, which themselves are translations of the Greek New Testament.

What is being studied as Old Church Slavonic is a purely literary language, and as Čaburgaev (1974, pp. 5-6) pertinently points out, it is important to remember that OCS is not the written variant of a spoken Slavic language at the time, but a language of the Church that spread in this capacity and acquired local traits in the different regions of the Slavic world where it was employed. By the time the redactions present in the PROIEL corpus (i.e. the manuscripts described below) were written down, different local traits can be expected from the manuscripts. Still, Čaburgaev, among others, considers this language to be close enough to the first translations from the 9th century to be able to call it Old Church Slavonic. In the study of OCS a specific set of texts, assumed to reflect a language not too different from reconstructed Common Slavic³², is traditionally referred to as "canonical". The inclusion into this canon is mainly based on spelling and morphological traits: more or less consistent usage of the reduced vowels referred to as *jers*, presence of nasal vowels and the vowel *ě*, as well as root aorists, among other things (Lunt, 2001, p. 6). The texts present in the PROIEL corpus – *Codex Marianus* and parts of *Codex Suprasliensis* – are both canonical texts. They have been annotated for morphology, syntax and information status. This is further explained in chapter 4.2.

The *Codex Marianus* (CM) is believed to be among the oldest existing OCS manuscripts, dating back to the early 11th century. It's a tetraevangelia and was originally written in Glagolitic. Flier (1974, p. 43) claims that CM is the text that comes closest to *the idealized grammatical norm*³³ of OCS. Its place of origin is believed to be Macedonia (Čaburgaev, 1974, p. 42) and it contains 174 folia, from Matthew 5:23 to John 21:7 (with John 1:1-1:22 and 18:14-18:28 missing³⁴). The entire CM is present in the PROIEL corpus and is fully annotated, but the IS annotation comes from the Greek text (explained in chapter 4.2).

The *Codex Suprasliensis* is the largest OCS manuscript, containing a total of 285 folia, of which a part is currently present in the corpus. This manuscript, which is a so-called *menaeum*, contains saints' lives for the month of March. It was intended for daily reading and was written in Cyrillic (Lunt, 2001,

³² This term is used to refer to the latter stage of Proto-Slavic, i.e. an undocumented language. Lunt (1989, p. 8) also says that Middle Common Slavic was very close to actual OCS.

³³ Since all manuscripts show certain regional variations, Flier operates with an idealized grammatical norm based on what should be the "best translations" of the original Greek. He takes into account Greek calques, innovations and increase in number of translations, which should hypothetically improve the competence of translators and bring forth more accurate translations, representing the actual content and not only copying forms and word order. See Flier 1974, pp. 31-51.

³⁴ <http://www-bcf.usc.edu/~pancheva/ParsedCorpusList.html> (01.10.2012)

p. 9). The manuscript, which is dated to the 11th century, shows language that is younger than that in CM, so it would not be unexpected to find some differences between CM and CS in the further analysis of the behavior of adjectives and participles. CS has only been partially annotated for IS, and in the analysis only these IS-tagged texts³⁵ will be included.

The IS-tagged sources used in this work comprise a total of 3421³⁶ adjectives and participles, both attributive and predicative, including nominalizations. After all necessary sorting and eliminating, which will be further explained in chapter 4.3.3, a total of 827 attributes and 911 nominalizations remain (excluding appositions and vocatives which are treated separately).

Table 4 below contains an overview of the distribution of adjectives and participles in the corpus (with percentages of the total column within parenthesis). These are the numbers before any exclusions have been carried out. As is seen in table 4, the great majority of the adjectival forms used in this thesis come from CM.

Table 4 The original distribution of adjectives/participles in the texts present in the corpus, prior to any exclusions

| Manuscript | Entire corpus | | From texts annotated for IS | |
|---------------------|---------------|--------------|-----------------------------|--------------|
| | Adjectives | Participles | Adjectives | Participles |
| Codex Marianus | 2580 (69.6%) | 3117 (70.6%) | 2580 (91.5%) | 3117 (92.2%) |
| Codex Suprasliensis | 1127 (34.4%) | 1300 (29.4%) | 241 (8.5%) | 263 (7.8%) |
| Total | 3707 (100%) | 4417 (100%) | 2821 (100%) | 3380 (100%) |

In addition to the OCS sources, the Greek New Testament³⁷ originals of the Gospel texts are also used for comparison.

4.2. PROIEL corpus annotation

As previously mentioned, the PROIEL corpus annotation includes morphology, syntax and information status. The morphological annotation needs not be further mentioned, except to say that each token is annotated for part of speech, tense, case, number, strength³⁸, etc. The syntactic annotation is based on dependency grammar; for more information on this, see the “PROIEL

³⁵ For this reason, the *Codex Zographensis* is not part of this study. Parts of it are present in the corpus, but lack IS annotation.

³⁶ This number for the entire corpus, regardless of IS-tagging, is 8325 as of April 2013.

³⁷ The Tischendorf Edition is the exact version used in the corpus. It is considered to be so close to the actual translations that there shouldn't be many discrepancies.

³⁸ This is the term used to refer to short (strong) and long (weak) form.

Guidelines for Annotation”³⁹ documentation. The information presented here on IS-tagging comes from “Guidelines for annotation of givenness”⁴⁰ and from an as of yet unpublished article by Haug, Eckhoff & Welo entitled “The theoretical foundations of givenness”.

The PROIEL corpus employs a range of IS-tags to mark what status the various referents of a text have, indicating what role they play in the narrative. Most of this annotation has been done in the Greek part of the corpus, and then, in extracting the data from the corpus into a dataset, Greek IS analysis is transferred using token alignments to match up with the OCS translation. The token alignment is carried out automatically, and then checked manually and corrected. Accuracy should therefore be high, especially as long as the original and the translation has the same syntactic structure. A consequence of this is that the *graphic* representation of the IS-tagging is only visible in the Greek texts, which must be consulted in order to see how referents are linked in chains. The exception is the collection of texts from Codex Suprasliensis, which have been tagged directly in the OCS version. The transfer of IS-tags causes certain problems that will be discussed at the end of this chapter. The following IS-tags are employed: OLD, OLD-INACT(ive), ACC(essible)-INF(erable), ACC(essible) -GEN(erally known), ACC(essible)-SIT(uational), NEW, ANCHORED (referred to as NEW-anchored), NON-SPEC(ific), NON-SPEC(ific)-OLD, NON-SPEC(ific)-INF(erable), KIND, QUANT(ified). They can be categorized as displayed in table 5:

Table 5 Tags in the PROIEL corpus (from the forthcoming Haug et al., p. 7)

| Context | Specific tag | Non-specific tag |
|---------------------|------------------|------------------|
| Discourse | OLD (+OLD-inact) | NONSPEC-OLD |
| Scenario | ACC-INF | NONSPEC-INF |
| Encyclopedic | ACC-GEN | |
| Situation | ACC-SIT | |
| | NEW | NONSPEC |
| | ANCHORED | |
| | KIND | QUANT |

Texts are annotated for information status using these tags from the *hearer's* point of view and only NPs that introduce discourse referents are annotated (forthcoming Haug et al., p. 5). Each tag will now be presented shortly. For more examples and a more exhaustive presentation of the annotation scheme, see the PROIEL-document “Guidelines for annotation of givenness”.

³⁹ http://folk.uio.no/daghaug/syntactic_guidelines.pdf (2013-05-01)

⁴⁰ http://folk.uio.no/daghaug/info_guidelines.pdf (2013-05-01)

4.2.1. NEW and OLD

Very simplified, something known to the hearer or which has an antecedent in the text is OLD and something unknown is NEW. For the NEW tag to be employed, a referent that cannot be inferred from the preceding discourse has to be introduced. Referents that have been previously mentioned are marked OLD and referred back to their latest mention. If this latest mention is farther away than 13 sentences, it is marked OLD-inactive. However, the OLD tag does not only pick up subsequent mentions of referents introduced as NEW. ACC-GEN and KIND that are picked up within 13 sentences are also tagged OLD. This distinction will be further mentioned in the analysis chapter.

NEW is, however, a tag that is less common than one could believe, and not all referents (even if they are specific) that are mentioned for the first time get this tag. In a corpus study from 1990 dealing with co-referentiality and definiteness in a variety of Swedish texts, Fraurud (1990, p. 401) has shown that indefinite NPs only represent 34.8% of the total amount of initial mention NPs, and that definite NPs stand for 27%. Definite form for first time mentions can be explained by several factors, such as that they are unique entities (see section 4.2.3), inferable from some other referent (see sections 4.2.4 and 4.2.5), or generic (see section 4.2.6). Definite first mentions are also found in the narrative technique referred to as *in medias res*. This is something that we see a couple of times in the Greek and OCS texts in parables, where a referent appears in the definite form “out of the blue”.

4.2.2. ACC-SIT

In dialogues, information disclosed by the participants represents a new universe. Referents may therefore be OLD to the reader who has the benefit of being familiar with the entire context, and NEW to a participant in a dialogue who has just entered “the stage”. The tag ACC-SIT is therefore reserved for referents referred to in direct speech that are accessible through situational cues (see forthcoming Haug et al., p. 6 and page 7 of the Guidelines). The example below illustrates this with a reference to people standing around the speaker, i.e. who are identifiable relative to his location.

- (19) ...нъ народа (ACC-SIT) ради **стоиштааго** окръсть рѣхъ. да вѣрж имжть ѣко ты ма
посъла. (CM, John 11:42, 42454)
...alla dia ton okhlon **ton periestōta** eipon, hina pisteusōsin hoti su me apesteilas
(22786)
... but I said this on account of the people **standing** around, that they may believe
that you sent me.

4.2.3. ACC-GEN

Naturally, certain things are known to the hearer without having been previously introduced. Since we cannot read the minds of the intended audience of these texts we assume that they are idealized 1st century Hellenized Jews, and generalize based on this, assuming that they have certain encyclopedic knowledge of the world. Referents considered “generally known” are tagged ACC-GEN

(accessible by general knowledge), picked up in subsequent mention as OLD. If the latest mention is farther away than 13 sentences, it is marked ACC-GEN anew. Example (20) contains six examples of ACC-GEN, one with an attribute.

- (20) Глѣже вамъ ѣко мнози отъ вѣстоу (ACC-GEN) и западѣ (ACC-GEN) приидѣтъ. ꙗко възлѣгѣтъ съ авраамомъ (ACC-GEN) и исакомъ (ACC-GEN). ꙗко и ѡвоу (ACC-GEN) въ црствіи (ACC-GEN) ѣбѣтъ. (CM, Matt 8:11, 38507)
 legō de hūmin hoti polloi apo anatolōn kai dusmōn hēxousin kai anaklithēsontai meta Abraam kai Isaak kai Iakōb en tē basileia tōn ouranōn (14919)
 And I say unto you, that many shall come from the east and the west, and shall sit down with Abraham, and Isaac, and Jacob, in the kingdom of heaven.

4.2.4. NEW-anchored

If a referent is formally NEW but not wholly unknown, it can be NEW-anchored. For all referents that are tagged NEW, the syntactic dependents for said referent are automatically checked for referents that are OLD or ACC-GEN⁴¹ when the data is extracted, as these are elements that can anchor the referent. Also, if the modifying adjective has an IS of its own that is OLD/ACC (as is the case for possessive adjectives), the NEW referent is considered to be anchored. Different behavior can be expected from these than from “purely” new referents. (21) is an example of a new referent modified by a possessive adjective that itself has the IS-tag ACC-GEN:

- (21) и оуона (NEW) црквнаѣ (LF ACC-GEN) раздѣра сѧ на двѣе съ выше до ниже. (CM, Mark 15:38, 37331)
 kai to katapetasma tou naou eskhithē eis duo apo anōthen heōs katō (7509)
 The curtain of the temple was torn in two from top to bottom.

In (21), the curtain is new in the narrative, but it is identifiable by association to the temple, and as such it is definite (OCS LF). Therefore, it is considered to be NEW-anchored in the IS annotation. However, this system does not function perfectly. Adverbials that contain anchoring elements do not trigger the ANCHORED tag, and some of the “unexpected LF” in the analysis chapter are made up of referents that should really belong in the NEW-anchored category. Other referents are erroneously anchored, as will be seen in chapters 5.1.2.3 and 5.2.1.3.

4.2.5. ACC-INF

ACC-INF is a tag that is used when a NP is inferable from some other known referent, cf. how a hand must necessarily (or hopefully) belong to someone or have belonged to someone. ACC-INF is also used when referring to the part-whole or whole-part type of relationships where we may first deal with a group of people and then with a particular member of that group (forthcoming Haug et al., p. 6). Assumptions concerning what kind of objects may be associated with people in the relevant time

⁴¹ The modification adding ACC-GEN to the anchoring script came about rather late in the process of writing this thesis, and therefore not all NEW referents with ACC-GEN adjectives have been picked up automatically. They are, however, commented on in chapter five and «moved» to the right category in the final analysis.

frame of the texts are made (i.e. what kind of belongings, relations, etc. people are expected to have), so that when the hermit in Paul the Simple closes a door, that door is referred back to the hermit:

- (22) и акы не рачи послоушати юго. затворивы **двьри** (ACC-INF) антонии. того ради не
въниде до трии дънии ни заходоу себѣ (CS, Paul the Simple, 58684)
And since he did not deign to listen to him, Anthony shut **the door** and because of
this did not enter for three days even into his latrine.⁴²

The ACC-INF tag is easy to over-use, and this is something that has been observed in the corpus annotation.

4.2.6. KIND

Generic referents such as “the sick” (when there is no specific group of sick people present in the narrative) are given the tag KIND and are picked up as OLD in subsequent mention. If the previous mention is farther away than 13 sentences, the tag KIND is used again. For more comments on generic referents, see chapter 2.2.3.

4.2.7. NON-SPEC

NPs that do not refer to any specific referent are tagged NON-SPEC, with the sub-tags NON-SPEC-INF and NON-SPEC-OLD. The NON-SPEC tag is used for *scopal* specificity and not for epistemic specificity. The latter, i.e. NPs that establish referents that are unknown⁴³ to the speaker and/or hearer are not counted as non-specific, they are tagged as NEW (forthcoming Haug et al., p. 16). The referents for non-specific elements in the PROIEL tagging scheme only exist within “certain embeddings”, outside of which they can no longer be referred to, including negation, modality and quantification (Ibid., p 10). Non-specific referents that are quantified are tagged QUANT (see 4.2.8).

The following example displays a non-specific referent with a modal expression.

- (23) како можетъ ѿлкъ **грѣшень**. сица знамениѣ творити (CM, John 9:16, 51970)
pōs dunatai anthrōpos **amartōlos** toiauta sēmeia poiein; (22593)
“How can **a sinner** perform such signs?”

4.2.8. QUANT

Quantified NPs are non-specific and given the tag QUANT (picked up as NON-SPEC-OLD in subsequent mention). What is considered to be quantified (as opposed to simply NON-SPEC) depends on formal properties. In Greek, headless relative clauses, nominalized participles with an

⁴² Translation from http://clover.slavic.pitt.edu:8080/exist/paul/data/paul_main.html (2012-08-30).

⁴³ Such referring expressions, where lack of identifiability in the referent is the key aspect in defining it as non-specific, are referred to as non-specific by for example Lyons (1999, p. 38), as in “There’s some book lying on the table”.

article and heads modified by overt quantifiers such as “every” are tagged QUANT (Guidelines for annotation of givenness, p. 9). In sentence (24) there is an example of QUANT (and the rare NON-SPEC-INF-tag⁴⁴).

- (24) **вськъ ѿлвкъ** (QUANT) прѣжде **доброє вино** (NON-SPEC-INF) полагаатъ. ꙗкогда оупиѣтъ сѧ тогда тачѣе (CM, John 2:10, 41683)
pas anthrōpos prōton **ton kalon oinon** tithēsīn, kai hotan methusthōsin ton elassō (21978)
Everyone brings out **the choice wine** first and then the cheaper wine after the guests have had too much to drink.

“Everyone” in (24), which is literally “every man” in both Greek and OCS, is a quantified expression and gets the tag QUANT. The wine referred to is not a particular wine, but good wine in general, and therefore has no referent outside of the context, and is non-specific.

4.2.9. Challenges related to the transfer of IS-annotation from Greek

In general, the transfer of IS status from Greek to OCS is unproblematic due to the token alignment. As long as the text has been correctly translated, the same referents should be NEW, OLD, etc. regardless of language. However, there is a certain number of “mistakes” concerning the distribution of KIND and NON-SPEC in the OCS part of the corpus, since generics and non-specific NPs are very similar. As it is not entirely known how exactly Greek marks generics, it has more or less been assumed (in the annotation of the Greek texts) that for referents that could be either non-specific or generic, the ones modified by a definite article are generic. This distribution will then have been transferred to the OCS data as well, regardless of the use of LF/SF there. Therefore there may be some confusion in the OCS dataset regarding the KIND tag (or what is supposed to have this tag), and one cannot always take for granted that the translator has “correctly” interpreted what is supposedly a generic expression in Greek and rendered this in some uniform way in OCS. As Flier (1974, p. 139) points out, “The difference between a generalized noun and an indefinite noun is not so great as to cause problems in interpretation if they are used interchangeably.” They only cause problems for linguists trying to classify them. Because of these difficulties, Appendix 1 contains a rather large amount of NON-SPEC referents that have been “transferred” to the KIND category, but also referents tagged KIND that have been transferred to the NON-SPEC category. This means that in the final analysis, these are not counted as “unexpected” uses of SF/LF.

⁴⁴ The wine has been linked back to “every man”, meaning that this good wine is inferable from what is considered the hypothetical possessor of said hypothetical wine (i.e. its reference will vary with every man), which renders NON-SPEC-INF.

When transferring the IS-tag QUANT from Greek, difficulties occasionally occur. If the syntactic structures of Greek and OCS do not match up perfectly, other elements than those that can be tagged QUANT in Greek can get this tag in the OCS text.

4.2.10. Concluding remarks regarding IS-tags

Since the general hypothesis of this paper is that SF and LF in adjectives and participles is connected to the pragmatics of the phrase they are found in, and that LF should correspond to something that is identifiable, we should expect to see a systematic correspondence between information status data and morphological data in the corpus. We would expect both the ACC-tags and the OLD-tag to have the same grammatical expression, i.e. LF attributes (or LF nominalizations). Generics, that is the tag KIND, and QUANT, in the sense of “all” or “every”, are other categories that we would expect to find LF in. NEW and NON-SPEC are expected to show SF.

4.3. Treatment of data

4.3.1. Data processing

In the analysis of the PROIEL data, the statistics program R⁴⁵ has been used. With the help of R, it is easy to single out groups of words that appear in certain relations to each other (based on the syntactic analysis in the PROIEL corpus), sort and analyze them.

Together with a certain amount of information, all OCS attributive, appositive and nominalized adjectives and participles have been extracted from the PROIEL corpus. These have been marked for LF or SF (or Undecided) together with a wide range of other morphological and syntactic traits. The dataset contains such information for each adjective and participle as source (manuscript), number, gender, degree, relation to head word, word class, gender and number of the head, the original Greek reference, conjunct elements, information status, etc. Altogether, there are approximately 100 columns with information, although far from all of these are always relevant. For example, manipulating this dataset in R enables one to easily extract all LF attributive adjectives that are female plural, subjects and tagged NEW, together with Greek originals if needed. The data analyzed in chapter five has been separated into an attributive dataset and a nominalized dataset, and these are then analyzed according to their information status. To see if the results are significant or not, statistical tests are carried out whenever relevant and the p-value for each category is presented together with the percentages.

⁴⁵ <http://www.r-project.org/>

4.3.2. A first view of the IS/strength distribution

A simple search can be made in the corpus data to see what correspondence there is between LF and SF and IS among attributes and nominalizations. This search has not been adjusted in any way, and it contains the basic data from the corpus. The less common IS-tags have been left out.

Table 6 Distribution of SF/LF across the most common IS-tags in the PROIEL corpus

| IS-tag | Attributes | | Nominalizations | |
|----------|------------|-----|-----------------|-----|
| | SF | LF | SF | LF |
| ACC-GEN | 195 | 127 | 275 | 97 |
| KIND | 24 | 45 | 23 | 84 |
| NEW | 222 | 66 | 28 | 39 |
| ANCHORED | 131 | 87 | 18 | 44 |
| NON-SPEC | 150 | 44 | 63 | 61 |
| OLD | 106 | 263 | 187 | 163 |
| QUANT | 123 | 34 | 92 | 185 |
| TOTAL | 951 | 666 | 686 | 673 |

In the ACC-GEN category, there is actually a majority of SF, which is contrary to the hypothesis. Among generics, LF is predominant, whereas SF clearly dominates the NON-SPEC category for attributes at least. Among the NEW and OLD, the exception rate seems surprisingly high. These numbers are bound to become more nuanced as we investigate the various categories more closely in chapter five, after certain exclusions have been carried out below in chapter 4.3.3.

4.3.3. Exclusions

In order to avoid categories that show no fluctuation between SF and LF or where SF and LF are not used based on pragmatics, the amount of data that will be used for the final analysis needs to be narrowed down. If a certain category of words always shows LF and it is clear to us why this is so, this group of words can be excluded. This sub-chapter will deal with the data that has been excluded.

4.3.3.1. Non-varying suffixes and lemmas

First and foremost, all adjectives with the suffixes *ov*, *in*, *j* and *nj* (see table 3 on page 22) have been removed, since these always appear in the SF. Adjectives with uncertain strength have also been excluded, as has certain adjectives that can be both short and long, such as *božii*.

4.3.3.2. The quantifiers *mъnogъ* and *въsjěkъ*

These two quantifiers are discussed in the chapter on nominalizations with the IS-tag QUANT. They show forms that appear to be more pronominal than adjectival, and are mainly short. *Mъnogъ* and *въsjěkъ* occasionally show up in other IS-categories as well, but the majority of them are found

among the quantified referents. Their uniformity also explains the absence of an attributive QUANT chapter.

Table 7 SF/LF distribution for the quantifiers *тѣпогѣ* & *вѣсѣкѣ*

| | SF | LF |
|----------------------------------|-------------|----------|
| Attributive <i>тѣпогѣ</i> | 106 | 5 |
| Nominalized <i>тѣпогѣ</i> | 70 | 1 |
| Attributive <i>вѣсѣкѣ</i> | 69 | 1 |
| Nominalized <i>вѣсѣкѣ</i> | 20 | 0 |
| Total | 265 (97.4%) | 7 (2.6%) |

Adjectives that do not show the behavior of regular adjectives are excluded, these two lemmas included.

4.3.3.3. The ordinal numeral *drugъ*

Ordinal numerals in general appear to display a great deal of LF, something that makes sense considering they are uniques in a series. “The second” is unique in relation to “the third”, etc. The OCS ordinal numeral *drugъ* carries both the meaning “the second” and “another”, but seems to display LF regardless of meaning.

Table 8 SF/LF distribution for the numeral *drugъ*

| | SF | LF |
|--------------------------|----|----|
| Attributive <i>drugъ</i> | 1 | 18 |
| Nominalized <i>drugъ</i> | 4 | 49 |

The LF nominalized *drugъ* contain 8 NEW and 8 NON-SPEC tags (i.e. 30.2% unexpected LF). The LF attributive ones contain 4 NEW and 1 NON-SPEC (i.e. 26.3% unexpected LF). On these grounds, *drugъ* is excluded from the final analysis.

4.3.3.4. The present participle *sy*

The m/n NA⁴⁶ present participle of *byti*, “to be”, *sy*, “being”, shows vacillation between SF and LF, and often shows up as an exception in the analysis chapter. There are only 7 examples of this participle, but the SF/LF distribution looks as follows:

⁴⁶ Nominative-Accusative

Table 9 Distribution of SF/LF for the present participle 'sy'

| | SF | LF |
|-------|----|----|
| NEW | 2 | 1 |
| QUANT | 1 | 1 |
| OLD | 1 | 1 |

Sy seems to behave rather erratically, possibly because it is usually adverbial. In chapter 5 this will only be mentioned briefly. Even though this lemma is *not* excluded from the dataset, it will not be counted as erroneous when it shows up with unexpected forms. When it appears in the LF, forms vary: *syi*, *soi*, *sęi*.

4.3.3.5. Coordinated attributes

For coordinated attributes in OCS only the first needs necessarily appear in LF to mark that something is identifiable (Flier, 1974, p. 50), something that is confirmed by PROIEL data. This entails that it is not necessarily the number of LF occurrences that is of interest, but rather the number of NPs modified by at least *one* LF attribute. Coordinated adjectives/participles that are not first conjuncts have therefore been excluded, since in the very great majority of coordinated LF, it is the first conjunct that is marked as long.

4.3.3.6. Predicative expressions

In the PROIEL corpus, predicative expressions are explicitly tagged as predicative complements to their head words. In modern Russian, predicative adjectives – i.e. “the man is **short**” – *mainly* make use of nominative SF adjectives or LF instrumentals (particularly in the past tense with an explicit copula), whereas LF is found in attributive and appositive adjectives (Mathiassen, 2009, pp. 93-94). The pattern we find in Russian is the one we also seem to *partially* find in the OCS of the manuscripts examined in this work, notably that predicative adjectives and participles display SF. Since predicative adjectives behave in a uniform manner, they will only be shortly mentioned here to see whether or not the corpus data confirm the general opinion on the matter.

Looking at the entire corpus (not only the IS-tagged part), we find 6.6% predicative participles in the LF, and 5.0% LF adjectives. That there is a certain amount of LF present here is explained by the fact that certain predicative expressions are identifying, as in “He is **Jesus of Nazareth**”, where both the grammatical subject and the predicative expression are referential. The amount of qualifying LF predicatives is small, but attributes to predicative heads that are referential are not excluded. All predicative headless adjectives and participles are excluded though, regardless of their referentiality. Example (25) illustrates a headless predicative SF.

- (25) отъвѣщаша имъ фарисѣи. еда и вы **прѣльштени** бысте. (CM, John 7:47, 42101)
 apekrithēsan autois hoi farisaioi mē kai humeies peplanēsthe; (22404)
 “You mean he has deceived you also?” the Pharisees retorted.
 OCS literally: answered them (DAT) Pharisees: “so and you **tempted** were?”

4.3.3.7. Conjunct participles

Conjunct participles, i.e. adverbials that modify the matrix verb, do not modify nouns. As such, we would expect them to all be in the SF. The amount of LF present is so small, 1.0%, that it does not have to be further commented upon. All conjunct participles are excluded.

4.3.3.8. Absolute datives

Absolute datives are adverbial temporal (or causal) constructions and have been excluded as they are of no interest for this thesis. The amount of absolute datives in the corpus is 184, of which 179, or 97.3%, are short.

4.3.3.9. Cooccurrences with demonstratives

If LF is a marker of definiteness it does seem reasonable that it should appear on attributes modifying clearly definite NPs, i.e. those that are modified by demonstratives. (25) shows a LF attribute (*ljutъjъ*) cooccurring with a demonstrative (*тъ*).

- (26) шедъ же **лютъи** слонуга **тъ**. ѿ вѣлѣзъ въ темницѣ повелѣ ѿмъ ѿзлѣсти (CS, Vita of the 42 Martyrs of Amorion, 83787)
That wicked servant having left and gone into the prison ordered them to leave.

Flier (1974, p. 74) says that this double marking of definiteness is a frequent phenomenon, even though according to him this is a case of “redundant marking” (see more on this in chapter 2.2.1). Borodič (1963, p. 164) states that there are *no* occurrences of SF attributes together with demonstratives in the Evangelic texts, making this seem like an absolute rule rather than a tendency.

The amount of NPs in the corpus that are modified by a demonstrative and an attribute is not very great. Some of the adjectives and participles found together with demonstratives are nominalizations, as illustrated by example (27).

- (27) Онъ же възрѣвъ на нѣ рече. что оубо естъ **писаное** се. (Luke 20:17, CM, 41245)
 ho de emblepsas autois eipen ti oun estin **to gegrammenon** touto (21516)
 Jesus looked directly at them and asked, “Then what is the meaning of **that which is written**”
 Greek/OCS literally: **the written** this

Altogether, we find 35 sentences, of which 15 contain nominalizations. Rather surprisingly, we find extremely few occurrences of *тъ* as demonstrative pronoun. In this sample, we have the following distribution:

Table 10 Distribution of demonstratives cooccurring with adjectives or participles

| Demonstrative | w. nominalization | w. attribute |
|---------------|-------------------|--------------|
| <i>sb</i> | 14 | 16 |
| <i>tb</i> | 1 | 2 |
| <i>onb</i> | 0 | 1 |
| <i>takb</i> | 0 | 1 |
| Total | 15 | 20 |

Sb, “this”, is clearly the most frequent demonstrative. The Greek originals (when there are any) predominantly show the combination article + demonstrative.

Table 11 Correspondence between Greek and OCS demonstratives

| OCS demonstratives with Greek originals | |
|---|--------------|
| sb | tb |
| - houtos: 24 | - ekeinōn: 1 |
| | - houtos: 1 |
| takb | |
| - toioutos: 1 | |

There is one sentence where the Greek lacks a definite article, and another one that has been so differently translated into OCS that the constructions do not match at all⁴⁷. Those sentences put aside, the pattern is rather clear. Of the total 35 adjectives/participles, only one is short: “the Son of Man”, *synъ člověčьskъ*, which is a special case.

- (28) кто **съ** есть ѡнь ѡлвчскъ. (CM, John 12:34, 42523)
tis estin **houtos** ho huios tou anthrōpou; (22855)
Who is **this** son of man?
Vad är väl **detta** för en **Människoson**?
Qui est **ce** Fils de l’homme?

93.8% of the total amount of “Son of Man” occurrences in the corpus are found in the LF, making (28) an exception. The Swedish translation, which differs radically from all the other, even uses an indefinite article, illustrating how this is possibly not a normal reference to the Son of Man, that is, to Jesus (even though the concept can also refer to “mankind”). Another possibility is that the concept is so established that the SF/LF distinction is not necessary, as both Flier and Borodič argue for very well-known referents. Considering how often this expression displays LF, this seems unlikely.

⁴⁷ Sentences IDs in question (OCS/Greek): CM, Mark 4:33: 36549/ 6694 and CM, Mark 1:27: 36396/6538.

The IS analysis in this entire group shows the expected ACC-GEN, OLD, etc., with very few exceptions. Cooccurrences with demonstratives are excluded since they show motivated use of LF.

4.3.3.10. Cooccurrences with possessives

In this section, the cooccurrence of three different possessive structures will be examined: possessive pronouns, possessive adjectives and possessive datives or genitives. Languages vary as to whether NPs with possessives have overt definiteness marking. This depends on the classification of the language as an AG or a DG language (see chapter 2.2.2). For an AG language, a definite article (in languages that have them), must be present to get a definite interpretation (Lyons, 1999, p. 130). What camp OCS belongs in is still uncertain, as it seems to show both types of behavior. Based on previous research, LF is generally expected with possessives, but perhaps not with all three constructions, as will be seen below. Unfortunately, some categories contain very few examples.

Just like with the case of cooccurring demonstratives, cooccurring possessive pronouns are not overly frequent. We would expect such phrases to denote known objects and individuals, at least for the speaker; “*our* heavenly father” may very well be fully unknown to a group of pagans.

A search in the IS-annotated corpus gives no more than 42 (28 attributive and 14 nominalized) instances with possessive pronouns of the kind *moi*, *tvoi*, *naš*, of which 3 (presented in examples (30)-(32) below) cooccur with short adjectives/participles. Example (29) **Error! Reference source not found.** illustrates a LF with a possessive pronoun, which we assume to be the norm for this group.

- (29) въроуієвъ бо богоу јако побѣжденъ бѣдешѣ нама. јакоже ѡ дїаволѣ побѣжденъ бысть. силѣ подаѣшѣтоу нама х҃соу. ѡ побѣдаѣшѣтоу **твоа проныривѣа кѣзни**. (CS, Vita of Paul and Juliana, 81208)
For we believe in God that you will be conquered by us, as the Devil was also conquered, seeing as Jesus gives us power and conquers **your lowly intrigues**.
- (30) и аште **десна твоѣ рѣка** съблажнаатъ та оуѣци ѡ. и врѣзи отъ себе (Matt 5:30, CM, 38383)
kai ei **hē dexia sou kheir** skandalizei se, ekkoupson autēn kai bale apo sou (14797)
And if **your right hand** causes you to stumble, cut it off and throw it away.
- (31) Тако и **отець мои ѡбскъ** сътворитъ вамъ... (Matt 18:35, CM, 39129)
houtōs kai **ho patēr mou ho ouranios** poiēsei humin (15546)
“This is how **my heavenly Father** will treat each of you....”
- (32) ѡ съберѣ тоу вѣсь жита моѣ и **добро мое**. (Luke 12.18, CM, 40751)
kai sunaxō ekei panta ta genēmata moi kai **ta agatha mou** (21000)
And there will I bestow all my grain and **my goods**.

Example (30) shows an exception to the rule that absolute opposites appear in the LF, since they represent a unique part of a whole. This must be deemed erroneous, and Greek is not to blame since the original is arthrous. Flier (1974, p. 127) states that *desnъ* always appears in the LF, with a couple

of exceptions, including the one displayed in example (30). This “right hand” has been given the IS-tag NEW in the analysis, but it is also anchored since the hand belongs to the addressee.

The SF in example (31) also seems erroneous. Possibly the abbreviated form of the adjective has some effect on LF being left out. Of the 9 examples of “heavenly father” in the entire corpus, 1 is short, and of the 57 occurrences of *nebesъskъ*, a mere 2 are short, the other occurrence being a nominalization; “heavenly things”. Since *nebesъskъ* will always refer to something connected to heaven, it’s head word is also always anchored⁴⁸. In example (32) we find a nominalization; “all my goods”, and LF is expected here as well, since the quantifier “all” renders the referent definite.

For the attributive group, the IS-statuses are all OLD/ACC-GEN/ANCHORED, with one odd NON-SPEC tag for a LF occurrence of “our daily bread”. In the nominalized group, we find five NEW and one NON-SPEC. Four of these are “your neighbor”, as in the example below, and all 6 are arthrous in Greek. The neighbor is hypothetical, but also definite (*the* unique neighbor of the addressee) in each specific case. These should be NON-SPEC-anchored or NEW-anchored.

- (33) възлюбиши **искрънѣго** **своего**. ꙗ възненавидиши врага своего. (CM, Matt 5:43, 50676)
 agapēseis **ton plēsion sou** kai misēseis ton ekhthron sou (47698)
 ‘You shall love **your neighbor** and hate your enemy.’

In this small group of SF adjectives with possessive pronouns we only find seemingly erroneous SF, and there is little doubt that adjectives/participles cooccurring with possessive pronouns display LF.

Cooccurrences with possessive adjectives constitute a very small group (6), and therefore it is difficult to say anything about them. Several of the occurrences display SF and have the IS-tag NEW, rendering meanings such as “one of X’s Y”, as in the qualifying (34), or are justifiably long, as in (35), where the LF participle “saying”⁴⁹ could either be viewed as marking the definiteness of the prophecy singled out by the possessor Isaiah, or, if Isaiah had many various prophecies, as a sort of cataphoric marker of definiteness, “the one saying the following...”.

- (34) ...**христовъ** же рабъ зѣло **простъ** сы. (CS, Paul the Simple, 58471)
 ...being a very **simple** servant of Christ’s.
 (35) и събываеъ сѧ имъ пророчѣсто **исаино** **ѣлжцеѣ** (CM, Matt 13:14, 50826)
 kai anaplēroutai autois hē profēteia **Hēsaïou hē legousa** (15207)
 In them is fulfilled the prophecy of Isaiah:
 Literally: and is fulfilled to them a prophecy of Isaiah, the one saying...

⁴⁸ In the modern sense of the word, “heavenly” may of course have other meanings, but in biblical texts we can assume that it is related to the one and only “Heaven”.

⁴⁹ This particular participle will appear several times in chapter 5 as SF where LF is expected. This indicates that it is probably more adverbial in nature, and only occasionally ends up tagged as an attribute when an appropriate verb is lacking.

The third sub-group is that of possessive nouns and pronouns in the genitive/dative cooccurring with adjectives or participles (29 occurrences). When the cooccurring possessive is a pronoun, LF is consistently found with one exception; when it is a noun, there is 33.3% SF, indicating more AG-like behavior – but this group is altogether very small. These SF instances present sentences similar to the one above (a simple servant of Christ’s). There is only one SF that should display LF (36), in which the prophets are definite (by inclusiveness in “all”), and their blood should also be definite:

- (36) да мьститъ сѧ крѣвь вьсѣхъ пѣркѣ. проливаема отъ съложениѣ вьсего мира. отъ
 рода сего. (CM, Luke 11:50, 40717)
 hina ekzētēthē to haima pantōn **tōn profētōn to ekkhunnonenon** apo katabolēs
 kosmou apo tēs geneas tautēs (20964)
 Therefore this generation will be held responsible for the blood *of* all **the prophets**
that has been shed since the beginning of the world.

The results from this subset show that LF is usually used when the referent cooccurs with a possessive pronoun. With the other groups, SF is just as possible when it is justified, i.e. when the possessive element does not in any way necessarily render the referent definite. LF is not arbitrarily used in the presence of any possessive element; its presence is instead motivated by the semantics of the phrase. Since these possessive instances have been examined here, they will not be included in the analysis in chapter 5 and are excluded from the dataset.

4.4. Concluding remarks

This chapter has presented the sources and the method used for this thesis. We have gone through the IS-tags that will now be examined for occurrences of SF/LF in the analysis chapter. Some data has been excluded based on morphological, syntactic, lexical or semantic/pragmatic factors, and the material that is now left should be easier to deal with and should present less repeated oddities.

5. Analysis – Attributes, nominalizations, appositions and vocatives

In the dataset that remains after all the exclusions explained above, we should find purely regular attributes and nominalizations whose SF/LF is not affected by other factors than the pragmatics of the phrase. The forms should only participate in signaling whether or not what is referred to is identifiable, generic, introduces a new referent, etc. Naturally, there is a certain amount of “errors”

present in all texts that can be explained in several ways. First and foremost, scribes surely made mistakes for no other reason than inattention or uncertainty. Secondly, when many of the texts were transcribed, the phonetics of the language were in the process of changing. The *jer*-shift⁵⁰, affecting the highly frequent phonemes represented by ѣ and ѥ, is perhaps the most noticeable change, and since certain letters then no longer represented phonemes, they could easily be confused or altogether forgotten. How the scribes worked can also help explain errors: they may have worked alone, copying from one manuscript into another, possibly dictating to themselves. Another possibility is that one person would read the text aloud, while several scribes copied it down (Flier, 1974, p. 50). In the second scenario, misplaced *jers* are very understandable if they were not even pronounced. If the written language was archaic, differing both in pronunciation and perhaps also in syntax from the spoken idiom, then mistakes are certainly to be expected. When SF or LF is found where it is not expected, these instances are referred to as “erroneous” or “unexpected”.

Throughout the analysis chapter, certain odd SF/LF will be deemed to not actually belong in the category they are found in. This is often explained by different interpretations in Greek and OCS (i.e. the OCS translator read the sentence differently), or by mistakes made by the annotators. When sentences have been “moved” to another category, the OCS sentences are found in Appendix 1. However, it is important to note that these “moved” sentences are not analyzed in whatever new group they could possibly fit better in. They are excluded from “unexpected SF/LF” but not actually “physically” added to another category. This way, in order to make my results more easily replicable, the results mirror what would be found by anyone else accessing the corpus data. When the terms “erroneous” or “unexpected” are used in reference to occurrences of SF or LF, these are cases that are *not* excluded or explained somehow. This is what counts in the “final analysis” numbers at the end of each subchapter, and is tested statistically.

The main goal of this thesis is to check the correspondence between SF/LF and various IS categories and see if SF/LF expresses definiteness, but the proposals of various linguists – presented in chapter three – will also be tested. These question have already been mentioned but will be repeated here:

- Are highly familiar referents marked by SF adjectives because using LF would be superfluous, as Čaburgaev indicates? This should become obvious in the ACC-GEN chapter (5.1.4).
- Is word order significant for definiteness? Is the pre-nominal position primarily occupied by indefinite modifiers, and the post-nominal by definite modifiers, as Dimitrova-Vulchanova and Vulchanov claim? Word order is examined in chapter 5.1.8.
- Is it true that LF only renders singulars generalized, not plurals, as Flier claims? Generics are examined in chapters 5.1.6 and 5.2.5.

⁵⁰ This is a process in which *jers* in “weak position” were lost, and changed phonetic value (which varied regionally) in the “strong position”. These vowels were distinct phonemes in the 9th century, but not by the time of the copying/translation of the texts in the PROIEL corpus (Lunt, 2001, p. 35).

- Is LF on adjectives derived from proper nouns so redundant that it almost never appears, as Flier seems to indicate? And do -ьsk adjectives derived from proper nouns display vacillation whereas they are always short when derived from common nouns, as Flier says, or is there a pragmatic reason for their use of SF/LF? This is to be examined straight away in chapter 5.1.1.

The analysis chapter will first deal with attributes and nominalizations. Appositions and vocatives are treated separately.

5.1. Attributes

In the group that has been narrowed down we find a total of 868 attributes (661 adjectives and 207 participles), all of which are annotated for IS. In the analysis of these, they will be dealt with separated into IS-tags. At the end of each sub-chapter, revised numbers, arrived at after a thorough examination of all the unexpected occurrences of either SF or LF, will be presented. This will give a more realistic view of the distribution of SF and LF than the numbers presented in chapter 4.3.2.

First of all in this chapter, the -ьsk adjectives will be looked more closely at, since they cause “problems” in several different categories. Then, the IS-tags NEW, OLD, ACC-GEN, NON-SPEC and KIND are examined. The two final sub-chapters deal with attributes to proper nouns and word order.

5.1.1. The -ьsk suffix

Since the -ьsk adjectives are seemingly more problematic than the other suffixed adjectives they deserve to be treated separately. It is obvious straight away from looking at the corpus data (see table 3 in chapter 3.5.1) that this suffix does not follow the OLD-LF and NEW-SF distinction we expect from regular adjectives, but it does not seem to be perceived as automatically making things known or definite either. If it were a definite marker on par with -ov, why is LF added to the -ьsk adjectives but not to the -ov adjectives? It seems to be something in-between the two categories. The -ьsk adjectives are not excluded from the discussion on attributes in the rest of chapter five, but they will most often not be discussed at length in the separate sub-chapters.

First of all, the -ьsk group of 248 attributive adjectives is rather homogenous regarding lemmas: *člověčьskъ* (92), *nebesьskъ* (46) and *ijudeiskъ* (25) represent 65.7% of all adjectives in this group. The remaining adjectives mainly denote places: *eleonьskъ* (9), *galileiskъ* (9) etc., peoples (nationalities) or known individuals: *fariseiskъ* (6), *cesarьskъ* (2), *elinьskъ* (2) and more general adjectives: *ljudьskъ* (5), *detьskъ* (1). A general note for all of these is that they follow their heads in 96.4% of the cases. For adjectives in general, this percentage is 77.4%, making this a statistically significant difference⁵¹. These adjectives will be separated into “unique entities”, “miscellaneous” and “places and

⁵¹ P-value < 0.0001, using a 2x2 contingency table and a Fisher’s exact test.

peoples”⁵². The placement of “miscellaneous” between the two other groups may seem a bit odd, but it is dealt with right after “unique entities” since both these categories display behavior similar to regular adjectives, whereas the group “places and peoples” does not.

5.1.1.1. Adjectives formed from unique entities

In this chapter we can remember Flier’s claim that -ьск adjectives show vacillation when derived from proper nouns, but show LF when derived from nouns referring to unique entities. The adjectives that have been formed from unique entities (144), with 126 (87.5%) LF and 18 (12.5%) SF, do show a majority of LF for OLD (93.0%) and ACC-GEN (100%), but also for NEW (77.8%) and NON-SPEC (66.7%), of which there are, however, not that many examples (9 for each). This is a category where the Greek originals display an article in 95.3% of the instances, with phrases such as “Son of Man”, “Kingdom of Heaven”, “Birds of heaven”, etc. They are also definite in English/French and the Scandinavian languages, and the suffixed adjectives seem to behave like regular adjectives for known referents. Since *člověčьskъ* (“of-Man”) is such a frequently occurring adjective (94 in total), it merits a separate comment. 83.0% (78) of occurrences show LF, but this does not necessarily have anything to do with the adjective itself: all of the heads of these LF adjectives have the IS-tag OLD or ACC-GEN, meaning that this adjective largely behaves like a regular adjective. For the set phrase “Son of Man” in particular, LF is at 93.8% (76), even though Flier (1974, pp. 107-108) claims that it was already so “fixed in terms of reference that it was taken to be representing Jesus with or without the long form”. Another very frequently occurring adjective is *nebesьskъ* (“of-Heaven”), of which 1 out of 56 is short, but more of these show the IS-tags NEW/NON-SPEC (mentioned below).

If these adjectives behave as regular adjectives, why are there many (7 out of 9) LF among the NEW? The NEW group displays several compounds for which it is doubtful whether they really are NEW, such as “clouds of Heaven”, “heights of Heaven”, “strength of Heaven”, “the kings of the Earth”, “judgment of Hell”. The modifying adjective in these cases has its own IS (i.e. Heaven, Earth) which is ACC-GEN. Their association to Heaven/Earth anchors them, rendering them identifiable. All NEW LF are such compounds, and it seems safe to classify them as NEW-anchored⁵³ or ACC-GEN altogether and rightfully LF. Some examples:

- (37) ...силы бо нѣбскыѧ подви҃гнѣтъ сѧ (CM, Luke 21:26, 51680)
 ...hai gar dunameis **tōn ouranōn** saleuthēsontai. (48574)
 ...for **the heavenly bodies** will be shaken.
- (38) посьла бо а҃г҃҃ла своѣго съ высотѣ **небесьскыихъ** съхранити ѿ. (CS, Vita of Paul and Juliana, 81752)
 For He sent his angel from **the heights of Heaven** to save her.

⁵² To see which lemmas are included in the various categories, see Appendix 2.

⁵³ That they are not picked up as such already in the data extraction is explained by technical reasons, and has to do with a late addition in the extraction code that turned ACC-GEN into an anchoring factor.

NON-SPEC that display LF show the same thing: “birds of Heaven”, which, since they are LF, do seem to be definite, and indeed are in all translations. These should be termed “NON-SPEC-anchored”. The only example that stands out among all the birds is the following in (39). All other usages of LF in this particular example correspond perfectly to ACC-GEN or OLD.

- (39) наречеть сѧ ѿнь бжїи. ѿ раждаѣтъ сѧ ѿтъ свѧтаѧго дѧа. ѿ ѿтъ непорочныѧ ѿ прѣчистыѧ дѣвы. **миръска’го** ради строѧ ѿ съпасения (CS, Vita of Paul and Juliana, 81150)

He calls himself “The Son of God” and is born from the Holy Spirit and of the immaculate and sinless virgin, for the sake of the order and salvation **of the world**.

SF seems just as plausible as LF here. Since LF is used throughout the sentence for known referents, the “order and salvation” referred to could very well have been interpreted as specific and referring by the translator (as it is anchored by “the world”), and the LF needs not necessarily be misplaced.

In the SF group, (12.5%, 18), IS-tags do not match up with SF use, as there are 6 OLD, 1 NEW-anchored and 5 KIND, for which we expect LF. The odd SF in the OLD category feature four of the five short “Son of Man”, and two other sentences with *člověčьskъ*. One of these could be read as non-specific (further discussed in chapter 5.1.3.2), but the rest are deemed erroneous. The generic SF are all instances of *člověčьskъ* as well, as is the NEW-anchored, which is an instance of “feet of man”, and should also be long.

All in all, for the “unique group”, there are a total of 12 “unexpected” SF, that, after some sorting (carried out above), are reduced to 11: the OLD, NEW-anchored and KIND SF; for all of these the adjective in question is *člověčьskъ*. None of the seemingly unexpected LF (NEW and NON-SPEC) is actually erroneous, since all of the referents were actually anchored. The error rate in total therefore seems to be 11 out of 144, or 7.6%, all due to one single lemma. The unique -ьsk adjectives as a group can therefore not be said to behave erratically.

5.1.1.2. Miscellaneous

The part of the -ьsk group containing various adjectives, the “miscellaneous” (19, featuring 12 different lemmas, 4 SF and 15 LF) continue to display the same patterns as the “unique group”. What is tagged NEW and has LF (6) is really inferable from the situation or world knowledge, and is actually anchored. It is generally definite in the Greek original and in translations into other languages. One example is *starьci ljudoьstii*, “the elders” (appears 3 times), which is annotated NEW. The elders of a village are usually a definite unit of people, and other people know who they are. Another example, the “depths of the sea”, where “the sea” anchors “the depths”, is very similar in genre to the “heights of Heaven”, and can similarly not really be counted as new and unknown to anyone.

Example (40) shows a sentence where the arthrous tombs have been tagged as NEW in the Greek text. The prophets they are associated with are, however, tagged ACC-GEN. As such, this should really have the tag NEW-anchored, and LF is expected.

- (40) Горе вамъ кѣнижъници и фарисѣи и ѱпокрити. Ёко зиждете гробы (NEW) пророчьскыя (ACC-GEN) и красите раки праведныхъ (CM, Matt 23:29, 51056)
 ouai humin, grammateis kai farisaioi hupokritai, hoti oikodomeite tous tafous tōn profētōn kai kosmeite ta mnēmeia tōn dikaiōn (15859)
 Woe to you, teachers of the law and Pharisees, you hypocrites! You build tombs for the prophets and decorate the graves of the righteous.

Among the LF, there is one clearly erroneous use of LF for “millstone” where the millstone in question is indefinite and tagged as NON-SPEC. The rest are rightly long. Among the SF, there is an unexpected SF in (41), where the children are OLD. The other instances of SF are all NON-SPEC.

- (41) ...ѡбо и пси подѣ трапезоу ѣдатъ отъ кроупиць (ACC-INF) дѣтескъ. (CM, Mark 7:28, 36732)
 ...kai ta kunaria hupokatō tēs trapezes esthiousin apo tōn psikhiōn tōn paidiōn (6884)
 ...yet even the dogs under the table eat the children's crumbs.

This group, which should be what Flier refers to as “derived from common nouns” should according to him show SF only. They do not – but on the other hand they are also quite few. All in all, there are only 2 forms that are deemed erroneous.

5.1.1.3. Adjectives formed from places or peoples

The adjectives formed from places or peoples (85), of which some naturally are proper nouns, seem more erratic. 84.6% of the 39 ACC-GEN show SF, as do 71.4% of the 14 OLD, whereas the NEW (17 in total) display 82.4% SF. Overall, the amount of LF⁵⁴ is very small at 15.

10 NPs tagged as OLD are modified by SF adjectives. In this group we find repeated occurrences of “land”, “sea” or “city” plus an adjective. In languages with articles, such constructions sometimes show articles and sometimes don’t. Expressions such as *zemi sodomьsce* and *zemljā genisaretъskā* are translated as “for Sodom/Sodoma/le pays de Sodome” and “at Gennesaret/til Gennesaret/dans le pays de Génésareth”, i.e. only French uses the word “land” and an article⁵⁵. They may appear more as actual names or set phrases with appositive adjectives that are conventionalized, than as attributes. This is typical in Early Slavic for countries, religions and languages (Eckhoff, 2011, p. 23), and because of this they will not be considered erroneous. More such occurrences are mentioned

⁵⁴ Among the appositions in chapter 5.3.1, we will encounter 13 -ъsk appositions formed from places or peoples that are *all* long, following the IS of their heads (all well-known referents).

⁵⁵ This is a very frequent turn of phrase in French. For example “Le pays de Galles” is “Wales”, and many French regions have names like “Le pays de X”.

below in the ACC-GEN and NEW categories. Four OLD SF that are difficult to explain remain: “the officers of the Jews”, “the king of the Jews” (2 occurrences), which should show LF as they are linked back to previous mentions, and the following example in (42), which should display LF since it also has a very clear antecedent in Matt 16:6:

- (42) храните же са отъ кваса фарисеиска и садукееиска (СМ, Matt 16:11, 39002)
 prosekhetē de apo tēs zumēs tōn farisaiōn kai zaddoukaiōn (15417)
 But be on your guard against the yeast of the Pharisees and Sadducees.
 Jag sade ju att ni skulle akta er för **fariseernas** och sadduceernas surdeg.

Only the Scandinavian languages do not explicitly mark the yeast/leaven as definite due to the genitive construction which renders it definite, but -ъsk does not, unlike -ov, automatically entail a possessive⁵⁶ sense. However, that it is actually possessive (i.e. a possessive “of the Pharisees” as opposed to a qualifying “Pharisean”) here *could* explain the SF. In Matt 16:12, *отъ кваса хлебънааго*, with a regular adjective, displays LF. The adjective *ijudeiskъ*, “Jewish” or “of the Jews”, will continue to show an unusual amount SF; only 3 out of 25 occurrences display LF.

The generally accessible category (ACC-GEN) with 33 SF and only 6 LF shows many of the same things as the OLD category. First of all the “King of the Jews” stands for 7 SF occurrences, and another 8 *ijudeiskъ* are also short in this group, including *pasxa ijudeiska*, *prazdnik ijudeiskъ* and similar compounds that remind of example (42). To this group another *kvasa fariseiska* and *uchenie fariseiska* can be added. There are 17 constructions such as the ones mentioned for the OLD category with “land” + adjective. Example (43), which has an article in Greek, illustrates this.

- (43) и исхождааше къ нему въсь иудейска страна и ѳрлмне и крѣштаахъ са въси въ іорѣданцѣи рѣцѣ отъ него исповѣдаѣхъше грѣхы своѣа (СМ, Mark 1:5, 36364)
 kai exeporeueto pros auton pasa hē Ioudaia khōra kai hoi Ierosolymeitai pantes kai ebapizonto hup’ autou en tō Iordanē potamō exomologoumenoi tas hamartias autōn (6505)
 The whole **Judean countryside** and all the people of Jerusalem went out to him.
 Confessing their sins, they were baptized by him in the Jordan River.
 Och hela **Judeen** och alla Jerusalems invånare kom ut till honom och bekände sina synder och döptes av honom i floden Jordan.

In Swedish, the “Judean countryside”⁵⁷ is simply “Judeen”, a proper noun (cf. the previously mentioned Sodom). Furthermore, *въсь*, “all”, “the whole”, is usually accompanied by LF, but not in *въсь ijudeiska* in example (43). The 8 occurrences of *gore eleonъsce*, “Mount of Olives”, seem to behave the same way. There are 3 instances of -ъsk adjectives denoting a place found as LF on referents tagged NEW; “all the hill country of Judea”, “all the region around the Jordan” and “the Asian desert” (from CS). For all of these, LF seems justified and they are hardly NEW, especially the

⁵⁶ Cf. chapter 3.5.1 and the categorization of the -ov suffix as possessive.

⁵⁷ Also note that this reference to “country” + name denotes the *people*, not a geographical unit.

first two. They are not similar to proper nouns but denote larger (anchored) areas. The “Asian desert” in CS is the only use of this adjective in the corpus, making it difficult to say anything about it.

Other “odd” SF are found in various IS-categories, but they will not all be presented in detail here both in order to save space and because it would be very repetitive. The arguments for why something is erroneous or not erroneous are the same as for the examples presented above (i.e. mostly “King of the Jews” and various countries or seas), and the final amount of unexpected forms is found in the table in section 5.1.1.4.

5.1.1.4. Conclusion regarding -ḅsk adjectives

All in all, the attributive -ḅsk adjectives do seem to conform to the general rules dictating SF for unknown and new, LF for known and accessible *except* for places or peoples (and certain occurrences of *člověčḅsḳ*). These are often derived from proper nouns, which, according to Flier, display vacillation, something the PROIEL data seem to confirm. A lot of the SF among these, *together* with their head words, look like proper nouns or concepts, where the adjective possibly has merged with the word to such a degree as to form an unchangeable unit, and has lost its function as a marker of indefiniteness/definiteness. Or, they are to be viewed as automatically rendering their head word definite, like the -ov adjectives. The Greek shows a lot of articles for these instances, showing that OCS has not simply copied the Greek constructions. In the table below the “error” percentage for each category is presented, showing that especially the “Unique” category behaves very much like regular adjectives. The amount of “erroneous” forms in the “Peoples & Places” category would be much larger if all the actual places (which behave as set phrases) were taken into consideration.

Table 12 Error percentages for -ḅsk adjectives

| | Uniques | Miscellaneous | Peoples & Places |
|--------------------|------------|---------------|------------------|
| Erroneous forms | 11 (7.6%) | 2 (10.5%) | 23 (27.1%) |
| Total (both SF/LF) | 144 (100%) | 19 (100%) | 85 (100%) |

Table displaying the amount of erroneous forms (for which no explanation could be found) in the various categories of -ḅsk adjectives.

5.1.2. Short and long form in the NEW category for attributes

A certain amount of oddities are found in the NEW category due to the way the corpus is annotated, with an initial LF percentage of 36.3% (see table at the end of section 5.1.2.2). What is formally new in the narrative gets the tag NEW, but quite often, these things are far from unknown or unfamiliar as they are frequently “anchored” to some well-known or previously mentioned referent. The issue of the NEW-anchored tag was discussed in the chapter on PROIEL annotation. This is a reoccurring

explanation for odd use of SF and LF in the NEW category, as will be demonstrated in the subsequent sections, and the actual percentage of LF for NEW referents will be established as a mere 1.4%.

5.1.2.1. IS-tag NEW and short form attribute

Since we expect SF for new referents, there is no need to comment at length on the occurrences that behave as expected. There are 74 short adjectives/participles modifying heads tagged NEW, of which a majority are adjectives (58 vs. 16 participles), and only one is arthrous in Greek. This particular expression, “good soil/ground”, is an expression that appears several times:

- (44) ...а дпоуое паде на земи **добръ** (CM, Luke 8:8, 40388)
...kai eteron epesen eis tēn gēn **agathēn** (20618)
...and other fell on **good** ground.

This expression displays SF as well as LF, and it is questionable whether it is actually referential, or rather, if it has been interpreted as having a referent by the translator, who has not matched the Greek article with an OCS LF. Instead, in the OCS version, it seems to belong in the NON-SPEC group, focusing on good as opposed to poor soil rather than on *a* specific soil. In the corpus it appears 6 times in SF, as NEW, OLD and ACC-INF. It is arthrous in Greek in 4 out of 6 instances, but the OCS translators do not appear to have seen this as a referring expression in any of the instances.

Typical examples of a newly introduced referents modified by a SF attribute can be seen in examples (45) and (46).

- (45) бѣ же тоу стадо **свино** пасомо велие при горѣ. (CM, Mark 5:11, 36575)
ēn de ekei pros tō orei agelē **khōirōn** megalē boskomenē (6723)
Now a great herd **of pigs** was feeding there on the hillside.
- (46) и абиѣ изиде бѣсъ. и прѣложи сѧ въ змии **великъ**. (CS, Paul the Simple, 58728)
And the demon immediately went out [of him] and turned into a **great** dragon.

5.1.2.2. IS-tag NEW and long form attribute

In the NEW category among attributive adjectives and participles there are but a few odd instances of LF. At a first glance, the LF percentage in this category is very high at 38.1%. However, many of these LF are easily explained, and the actual number of erroneous LF is indeed much smaller at 1.4%, represented by *one* single LF from CS for which no obvious explanation can be found. 17 can be discarded on the basis that they are definite by association (they are anchored) or belong in the ACC-GEN/INF/SIT category (see Appendix 1). These are typically sentences such as “clouds of Heaven” and are all arthrous in Greek, and the adjectives themselves frequently have the IS-tag ACC-GEN.

In most cases where we find LF, it is usually as a marker of identifiability, signaling that the referent is not entirely unknown. A new referent *is* introduced, but this referent is identifiable by the hearer once some additional information has been provided in what corresponds to a restrictive relative

clause⁵⁸, meaning that these referents are, strictly speaking, not new, only formally so in the linear representation of the message in the narrative. A frequent combination is "X, called-(LF) Y" with the verb *naricati*, "to name, call" (9 occurrences), as in example (47). These kinds of sentences do not show erroneous use of LF.

- (47) и бысть ѿко приближи сѧ въ витьфаѡиѧ и витаниѧ. къ горѣ нарицаемѣи елеонѣ (CM, Luke 19:29, 41184)
 kai egeneto hōs ēngisen eis Bēthfagē kai Bēthanian pros to oros to kaloumenon elaiōn... (21451)
 As he approached Bethphage and Bethany at the hill **called** the Mount of Olives...
- (48) ...и при четвъртъи стражи **ноштъни**. приде къ нимъ по морю ходѧ. (CM, Mark 6:48, 36684)
 ...peri tetartēn fulakēn **tēs nuktor** erkhetai pros autous peripatōn epi tēs thalassēs (6835)
 ...and about **the fourth watch of the night** he cometh unto them, walking upon the sea, and would have passed by them.

Another recurring situation in which we find new referents with LF attributes is with ordinal numerals, represented by example (48) above where "the *fourth* watch"⁵⁹, i.e. the ordinal number (by virtue of being a unique part of a series), clearly triggers LF, even though the watch itself is formally new. Something similar is found in (49), where the Sabbath in question is new in the narrative, but, since it has its own name, it seems likely that it is a generally known Sabbath, i.e. ACC-GEN. The name itself specifies what Sabbath it is: the second after the first. As such, the annotation of it as NEW can be questioned, and this is not counted as an erroneous LF.

- (49) Бысть же въ суботѣ **второпръвѣи** (CM, Luke 6:1, 40221)
 Egeneto de en sabbatō **deuteroprōtō** diaporeuesthai auton dia sporimōn (20446)
 And it came to pass on the **second sabbath after the first**.

One example of a "dubious" NEW-tag is found below in (50), in the text describing the plot to kill Jesus. The use of LF here can be attributed to uniqueness, i.e. each village/city has a specific set of elders, as mentioned in chapter 5.1.1.2. All of the referents mentioned are accompanied by articles in Greek, and are definite in the Western translations, so one could assume that they are known individuals. This NEW-tag should either be NEW-anchored (i.e. anchored by "the people") or be altogether "moved" to the ACC-GEN group.

- (50) тъгда събѣрашѧ сѧ архиереи и кѣнижѣници и старѣици **людьстии**. на дворѣ архиереовѣ. нарицаемаго каиѣфа. (CM, Matt 26:3, 39580)
 tote sunēkhthēsan hoi arkhierēis kai hoi presbuteroi **tou laou** eis tēn aulēn tou arkhierēōs (47998)

⁵⁸ These often feature oblique additions, as in "the way leading *into the cave*" which introduces an element of uniqueness by association to a specific cave, as we can suppose that caves (often) have *one* entrance.

⁵⁹ This is an indication of time corresponding to 3-6 a.m.

Then the chief priests and the elders of the people gathered in the palace of the high priest.

The remaining LF uses, both from CS, are presented below. Example (51) seems like the only example that is actually erroneous, unless the “indispensable need” is a concept that is lost on “modern” readers. Most likely it is not; Antonius has simply not left his house for four days, waiting for Paul to leave him alone, and finally he goes out because he had “an indispensable need”. The example in (52), since it is in the LF, could lean towards being an established concept (ACC-GEN), something that could be translated as “abstemious bliss” (or, as in the translation below, “restrained grace”). Another interpretation is that this “restrained grace” is a referring expression, referring back to the one year of spent in solitude by Paul, and the translation should really be “having led *this* restrained grace”. In that case, this should have the tag OLD or ACC-INF.

- (51) Въ четвъртъи днь. **потръбная** ноужда ємоу бысть. и отврѣзъ изиде. и видѣвъ паки павѣла глагола ємоу... (CS, Paul the Simple, 58468)
On the fourth day he got an **indispensable need** and having opened (the door) went out, and after having seen Paul again he said to him...
- (52) и благодѣти съподоби ся о бѣсѣхъ. и болѣзнихъ вѣсѣхъ. оуправивъ до коньца **въздрѣжаньнжж** благодѣть (CS, Paul the Simple, 58692)
And he became worthy of grace concerning demons and all sorts of illnesses, having led a thoroughly **restrained grace**⁶⁰.

Both of the examples above show adjectives placed pre-nominally and are from a text that is younger than the other ones. Pre-nominal adjectives are a minority in the corpus at roughly 30%. Altogether, the NEW category displays a very small amount of inexplicable LF uses. If we exclude adjectival constructions corresponding to restrictive relative clauses and instances with *naricati*, the final numbers are the ones presented below. What is left in the NEW category is now what is *really* new and unknown, and not immediately accompanied by “enlightening information”.

Table 13 LF/SF distribution in the NEW category for attributive adjectives and participles

| | SF | LF |
|--------------------------|-------------|------------|
| From corpus data | 72 (63.7%) | 41 (36.3%) |
| Dubious cases to exclude | 0 | 40 |
| Final analysis | 72 (98.6%)* | 1 (1.4%) |

Table displaying the original amount of SF/LF in the NEW category and the final percentage arrived at after the sentences have been examined. Exclusions include referents that are ACC-GEN or NEW-anchored (see Appendix 1).

*p-value < 2.2e-16, 95% confidence interval: 92.6%-100.0% (using a binomial test)

⁶⁰ Translation from: http://clover.slavic.pitt.edu:8080/exist/paul/data/paul_main.html (2013-05-01)

5.1.2.3. Short and long form for NEW-anchored referents

The anchored category, with a total of 65 attributive adjectives and participles, displays 35.4% SF (23). To begin with, the percentage of SF would be smaller if *all* referents that are actually anchored were to be found in this group (i.e. those not picked up by “the system”), but that’s not the case. Many in this group do not actually belong here, and the actual SF percentage is quite similar to the other categories (OLD, ACC-GEN, etc.) where we expect LF. Only the SF occurrences will be examined here. The expected LF is represented by example (53), where the ACC-INF anchors the NEW referent.

- (53) не можеш бо ты сътрѣпти скръби (NEW) поустыи́нскихъ (ACC-INF) (CS, Paul the Simple, 58668)
For you cannot endure the rigors of the wilderness.⁶¹

Among the (unexpected) SF we find many that are not actually anchored, based on the following factors:

- NEW referents that are anchored by subsequent mentions referring back to themselves, where the subsequent mention is tagged OLD, as in example (55). There are 8 such sentences.
- NEW referents that are anchored by an ACC-GEN, ACC-INF etc. that do not actually anchor it, for example in a comparison, or where the anchoring factor is actually non-referential, as in (54), or in clearly introductory sentences (56). This last category contains various sorts of sentences that cannot all be presented here. All 9 are found in Appendix 2.

- (54) Пакы подобно естъ црство небское неводоу (NEW) въврженоу въ море (ACC-GEN) ꙗко отъ всѣкого рода избѣравъшу. (CM, Matt 13:47, 38841)
Palin homoia estin hē Basileia tōn ouranōn sagēnē blētheisē eis tēn thalassan kai ek pantos genous sunagagousē (15258)
Again, the kingdom of Heaven is like a net that was thrown into the sea and gathered fish of every kind.
- (55) Бѣ же тоу единѣ члвкъ (NEW). ѿ. ꙗко осмь лѣтъ имы въ недѣлѣ своемъ (OLD). (CM, John 5:5, 41853)
ēn de tis anthrōpos ekei triakonta kai oktō etē ekhōn en tē astheneia autou. (22149)
One man was there who **had** been an invalid for thirty-eight years.
OCS literally: there was there a man, thirty-eight years having in his illness.
- (56) ...къ дѣвѣ (NEW) обрченъ мжеви. емоуже има иосифъ. отъ домоу дѣвѣ (ACC-GEN). ꙗко има дѣвѣ мариѣ. (CM, Luke 1:27, 39947)
... pros parthenon emnēsteumenēn andri hō onoma Iōsēf ex oikou Daveid, kai to onoma tēs parthenou Mariam (20175)
...to a virgin betrothed to a man whose name was Joseph, of the house of David. And the virgin's name was Mary.
- (57) аминъ глѣхъ вамъ. ꙗко единъ (NEW) отъ васъ прѣдастъ мѣ. ѣды съ мѣнож (OLD) (CM, Mark 14:18, 37182)
amēn legō hūmin hoti eis ex humōn paradōsei me, **ho esthiōn** met emou (7354)
Truly, I say to you, one of you will betray me, one **who is eating** with me.

⁶¹ Translation from: http://clover.slavic.pitt.edu:8080/exist/paul/data/paul_main.html (2013-05-07)

The “sea” in example (54) does not in any way anchor the net, which, together with said sea, is clearly hypothetical and non-referential. The man in (55) is anchored by “his” in “his illness”. The virgin in (56) remains un-identified until the subsequent sentence, where we learn that her name is Mary, and SF is expected on “betrothed”. The anchoring factor is the mention of “house of David”. Some examples are more difficult though. In example (57), the man who will betray Jesus is one of those who ate with *him* (the anchoring factor), but several people did, and the one person is possibly not yet singled out. The Greek seems to single him out, but perhaps not the OCS translator, nor does the English, the Swedish or French. There are two such sentences where different readings give either an anchored or purely new meaning, but SF cannot be considered erroneous as these sentences *can* be read as non-specific/new. The final problematic sentence is a SF dual featuring “the feet of man”, which should probably be long. However, as was seen in chapter 5.1.1.1, the adjective *člověčeskъ* is erratic in its use of SF/LF. This will still be counted as an error.

Once all such sentences are removed, only four can be deemed erroneous, meaning that SF is actually at 8.7%. All of the wrongly anchored sentences belong in the NEW or NON-SPEC⁶² categories, and would “improve” the percentages there further. The full list of “wrongly anchored” referents can be found in Appendix 1. As an example of an unexpected SF, there is sentence (58).

- (58) блаженое чръво (NEW) **ношъше** тѧ (OLD). ꙗ съсъца ѣже еси съсалъ (CM, Luke 11:27, 40682)
makaria hē koilia **hē bastasasa** se kai mastoid hous ethēlasas (20929)
“Blessed is the womb **that bore** you, and the breasts at which you nursed!”

In the sentence above, the anchoring element is *tę*, “you” sg.acc, which refers back to Jesus, rendering the womb definite in its uniqueness: only one womb can have born Jesus.

Table 14 LF/SF distribution in the NEW-anchored category for attributive adjectives and participles

| | SF | LF |
|--------------------------|------------|-------------|
| From corpus data | 23 (35.4%) | 42 (64.6%) |
| Dubious cases to exclude | 19 | 0 |
| Final analysis | 4 (8.7%) | 42 (91.3%)* |

Table displaying the original amount of SF/LF in the NEW-anchored category and the final percentage arrived at after all the sentences have been examined. Exclusions mainly include wrongly anchored referents that are either purely NEW or NON-SPEC. Sentences are found in Appendix 1 and 2.

*p-value = 5.1e-09, 95% confidence interval: 79.2%-97.6% (using a binomial test)

⁶² Comparisons and hypothetical referents found in parables probably belong in the NON-SPEC category, rather than in the NEW category, as they do not establish a referent outside of a certain context.

5.1.3. Short and long form in the OLD category for attributes

The OLD category shows fewer unexpected combinations than the NEW category, with a LF percentage of 89.6% that confirms the hypothesis of LF for old and known referents before any extra investigations have been carried out. Once erroneous instances are excluded, the LF percentage will be at 95.1%. This group includes OLD-inact, i.e. referents picked up outside of the range of 13 sentences, but they are not dealt with separately.

5.1.3.1. IS-tag OLD and long form attribute

173 attributes display LF among the heads tagged as OLD, i.e. 89.6%, but as already mentioned, the real LF percentage is 95.1% once dubious OLD-tags have been removed (see 5.1.3.2.). Two examples of OLD LF with an anaphoric⁶³ link is found in (59), with “the guarding soldiers” referring back to “the soldiers” (which also refer back to a previous mention of the same soldiers that is NEW).

- (59) ѿ възвѣштено бысть аурилияноу ѿтъ **приставыѣнныхъ** войнъ все то бывъшеѹ.
[...]и пославъ ѿтведе **стрѣгштыя** войны ноштыѹ. (CS, Vita of Paul and Juliana,
81274 & 81278)

And Aurilian was told by the soldiers **that were placed there** all that had happened
[...] and having sent message he led away the guarding soldiers in the night.

Only 8 of the LF sentences that have Greek originals (143) are anarthrous, and as such OCS seems to follow Greek article use on heads closely. As a reminder of that accessible (ACC-GEN) referents that are picked up again in subsequent mentions in the narrative are tagged as OLD it can be of interest to have a look at the frequency of a couple of the lemmas represented in this category:

Table 15 Distribution of lemmas in the OLD long form category

| lemma | number | % |
|------------|--------|-------|
| člověčьskъ | 68 | 39.1% |
| nebesьskъ | 14 | 8.0% |
| světъ | 10 | 5.7% |
| Total | 92 | 52.9% |

What we are dealing with here is, obviously, a large amount of “Son of Man”, “Kingdom of Heaven” and “Holy Spirit”, where it is not a question of purely anaphoric definiteness marking of the sort “**A sick woman** was lying in the bed. **The sick woman** later died.” We are instead dealing with general cultural knowledge and referents that would be definite even as first mentions. Several other

⁶³ Here “anaphoric” is understood as referring back to a newly introduced referent, not to a generic or generally known referent that would display LF regardless of whether it was part of an anaphoric chain or not.

instances are generic referents picked up as OLD. All in all, purely anaphoric referents are not as frequent as one may think.

5.1.3.2. IS-tag OLD and short form attribute

There are 20 OLD SF. 3 are anarthrous and 14 arthrous in Greek (the 3 remaining sentences lack Greek originals). For the anarthrous sentences, the question is mainly why these are OLD to begin with if they lack articles in Greek. Below examples and exceptions will be presented, which will show that the “error” percentage for OLD is closer to 5.0%. When OLD tags that can be interpreted as NON-SPEC or for some other reason have a justified SF are excluded, 11 clearly OLD ones remain.

Among the NPs that are fully arthrous in Greek there are 6 instances of various heads modified by *člověčeskъ*, which were discussed in chapter 5.1.1.1 as well. Four of these are “Son of Man”, which should arguably be long since Jesus is a well-known individual, and for which use of SF will here be considered an error on behalf of the OCS translator. The one mention of the “heart of man” in (61) should probably also be long, but “paradosis of man” in (60) could be rightfully short if the scribe interpreted it as plural non-specific, as in the English translation:

- (60) о́ставше бо заповѣдь божіѣ. дрѣжите прѣданиѣ (pl) **члвчска** (pl). (CM, Mark 7:8, 36702)
afentes tēn entolēn tou theou krateite tēn paradosin (sg) **tōn anthrōpōn** (pl) (6853)
You ignore God’s commandment while holding on to rules **created by humans**...
OCS literally: “rules of-Man” or “rules human”
- (61) из оутръждѣ бо отъ срдца **члвчка** помышлениѣ исходятъ. (CM, Mark 7:21, 36721)
ek tēs kardias **tōn anthrōpōn** hoi dialogismoi hoi kakoi ekporevontai (6872)
It’s from the inside, from the **human heart**, that evil thoughts come.

Among the rest, some should display LF, as in examples (62) and (63), and some could just as well have been interpreted as non-specific by the translator, as in (64).

- (62) тожде же и разбоиника **распата** съ нимъ поношаасте. емоу (CM, Matt 27:44, 39782)
to d’auto kai hoi lēstai **hoi sustaurōthentes** sun autō ōneidizon auton (16206)
In the same way the rebels **who were crucified** with him also heaped insults on him.
- (63) оубоудѣше же сѧ видѣша славѣ его. и оба мжжа **стоѹща** съ нимъ. (CM, Luke 9:32, 40514)
diagrēgorēsantes de eidan tēn doxan autou kai tous duo andras **tous sunestōtas** autō (20752)
But they managed to stay awake and saw his glory as well as the two men **with him**.
OCS literally: two men standing with him

The rebels mentioned in (62) refer back to a previous mention of “two rebels” and is clearly anaphoric OLD. The SF here could be explained by this being a dual⁶⁴ form, of which there are two SF instances in the group of 20 OLD SF. The *total* amount of duals among all the attributes in the narrowed down dataset is 12 (1.4%), of which merely 2 are long. The second SF dual is found in sentence (63), with a participle corresponding to a restrictive relative clause, and LF is expected. However, as with example (62), it may be relevant that this is a dual form.

- (64) ...и въврѣжѣтъ ѿ въ пещь **огньнѣ** (CM, Matt 13:50, 38845)
 ...kai balousin autous eis tēn kaminon **tou puros** (15261)
 ...and throw them into **the blazing furnace**.

Example (64) is different in that “the blazing furnace” seems to have been interpreted as *any* blazing furnace by the translator. The important fact transmitted in this sentence is not that they will be thrown into a specific blazing furnace, but that they (the wicked), contrary to the righteous, will burn. The furnace may very well have been mentioned previously (and it has), but it does not necessarily play an important role as a referent in this sentence.

Another SF is found in “living water” in John 4:11, *voda živa*, (Greek *to hydôr to zôn*) which is a concept often referring to the teachings of Christ, listed with both short and long form as *voda živa(ja)* and *voda životьna(ja)* in *Frazeologičeskij slovarь staroslavjanskogo jazyka* (Šuležkova, 2011, p. 83). It appears both in LF and SF in the corpus, but the number of occurrences is very small: 3. Borodič (1963, p. 167) mentions this very example and argues that it should display LF since this is a second mention, and therefore refers back to the living water mentioned in John 4:10. In all translations it is definite, usually with a demonstrative. As such, SF seems erroneous in this particular example, but otherwise it does not seem to be unusual judging from examples that can be found in the above-mentioned dictionary from other sources, meaning that this is probably sufficiently conceptualized to not need any LF marking to signal that it ought to be identifiable.

Other exclusions feature the previously mentioned “good soil” (3 occurrences), and as has already been discussed, this concept could just as well be interpreted as non-specific and probably has been by the translator. All in all, 8 out of 14 SF in the group that is arthrous in Greek can be said to be errors, i.e. cases that should show LF in OCS. These include instances of “Son of Man” and duals.

In the anarthrous group and among those lacking Greek originals, SF use is rather easy to explain, and only one out of the 6 (presented below) will be deemed erroneous, meaning the rest can be excluded from the final statistics.

⁶⁴ According to Vlasto (1986, p. 113), duals disappeared early in the adjective, even in OCS. Therefore, it may not be surprising to find more erroneous forms among duals than among singulars and plurals.

- (65) дѣло бо **добро** сътвори въ мѣнь. (CM, Matt 26:10, 39588)
ergon gar **kalon** ēgasato eis eme (16010)
 She has done a beautiful thing for me.

Dobro in (65)⁶⁵ does refer back to a previous action, and as such it is OLD, but even so SF is not surprising if the emphasis is on good as opposed to evil, and not on creating a clear reference to a specific action. There is another example where the emphasized, contrastive element may explain SF use where LF is expected, presented below in (66).

- (66) видѣвъше вси рѣптаахъ ꙗ́коже. ꙗ́ко къ **грѣшньмъ** мѣжю вѣниде витать. (CM, Luke 19:7, 41149)
 idontes pantes diegonguzon, legontes hoti para **amartōlō** andri eisēlthen
 katalusai (21415)
 Everyone who saw this grumbled, saying, “He has gone to be the guest of a sinner.”
 OCS/Greek literally: sinful man

In (66), the sinful man in question has been mentioned previously, but SF can be justified if emphasis is on the idea that Jesus going to *a* sinful man, as opposed to an honorable one (as is the case in the English version). In this reading, the man should be tagged as non-specific. Another sentence where it is doubtful whether SF can be considered an error is the following, where both an indefinite and definite interpretation would give the same meaning:

- (67) и ꙗ́коже ѿлѣгоу имъ ѿстоумоу **соухъ** ръкъ (CM, Mark 3:3, 50264)
 kai legei tō anthrōpō tō **tēn xēran** kheira ekhonti (6612)
 He said to the man with the withered hand.

The withered hand is an identifying trait for the man, and “with a withered hand” or “with the withered hand” makes no difference for the reference of the man, who has been mentioned earlier together with said withered hand. The participle “having”, *imōštumu*, is in the LF, making the man definite, and as such, also marking the hand as definite may have seemed irrelevant.

- (68) ꙗ́коже исплънь дѣха **ста**. възврати сѧ отъ иорѣдана. (CM, Luke 4:1, 40094)
 lēsous de plērēs pneumatōs **agiou** hupestrepsen apo tou iordanou (20315)
 And Jesus, full of the **Holy Spirit**, returned from the Jordan.

The last SF is found in a mention of “the Holy Spirit” in (68), which should display LF, but lacks article in Greek. The heavily abbreviated form in *dcha sta* may explain the SF, something which is further discussed in the next chapter, or the fact that this refers to the *substance* of “holy spirit”. As such, at the most *one* of the OLD SF corresponding to Greek anarthrous NPs is necessarily erroneous. This is not surprising considering that the Greek lacks articles here as well, and the translator could then

⁶⁵ This very example also shows up in the ACC-INF category – where it probably belongs. See example (77) on page 62.

have interpreted these expressions as having non-specific referents, even though the presence of previous mentions has been established.

Among the three sentences lacking Greek originals, we find two mentions of God with a SF adjective, both from Codex Suprasliensis:

- (69) ник'тоже бо можааше свойма очима ба **бесплътна** видѣти (CS, Vita of Paul and Juliana, 81154)
For no one could with their own two eyes see the **bodiless** God/God, **the bodiless one**.
- (70) нъ простивъши ны яко **блага** бога служителъница. помоли сѧ хѣоу да подасть намъ видѣнїя (CS, Vita of Paul and Juliana, 80934)
And having forgiven us as a servant of the **good** God, pray to Christ for him to give us a sign!

These are the only 2 SF modifying God among the 13 in the narrowed down dataset (with IS-tags OLD or ACC-GEN). There is no lack of mentions of God in general in the corpus, but he is seemingly not very often accompanied by attributes. It does not seem to be *wholly* unusual for very well-known referents (which are often abbreviated) to be modified by SF⁶⁶, but the adjectives in question here are not heavily abbreviated, so this does not seem to be a plausible explanation for SF. (69) could be non-specific, rendering “for no one could with their own two eyes see a bodiless God”. If (69) were to be interpreted as appositive (“God, the bodiless one”) we would expect LF. In (70) the SF seems odd, and we could disregard altogether of *blaga* as an attribute to *boga*, and instead see it as a modifier of *služitel'nica* (“a good servant of God”). If that is the case, the “error” percentage of SF in the OLD category decreases even further to 3.9%.

The final sentence is somewhat peculiar in that it refers back to a previously mentioned wind. The OCS uses plural, the modern languages used in this thesis for comparison all use the singular, and many use predicative constructions:

- (71) вида же вѣтры **крѣпкы** оубоѡ сѧ. (CM, Matt 14:30, 38905)
But when Peter saw the strong wind, he became frightened.
Men när han såg huru **stark vinden** var, blev han förskräckt. (Predicative.)
Mais, voyant que le vent était **fort**, il eut peur. (Predicative.)

Possibly, the OCS expression here is predicative, as is the Swedish and French, or the translator simply did not see this as referring back to a specific wind, and saw no need to individualize it using the LF. Furthermore, the previous mention of the wind in Matt 14:24 is singular: *bě bo protivenъ větrъ*, “for the wind was contrary”. If the -y in *větry* in (71) is a slip of the pen (ѣ and ы not being terribly dissimilar), then *krěpky* is a perfectly fine singular LF. If indeed plural, LF would be -yě.

⁶⁶ Out of 27 OLD and ACC-GEN tagged instances of “Holy Spirit”, 5 display SF, all heavily abbreviated.

Table 16 SF/LF distribution in the OLD category for attributive adjectives and participles

| | SF | LF |
|--------------------------|------------|--------------|
| From corpus data | 20 (10.4%) | 173 (89.6%) |
| Dubious cases to exclude | 11 | 0 |
| Final analysis | 9 (4.9%) | 173 (95.1%)* |

Table displaying the original amount of SF/LF in the OLD category and the final percentage arrived at after all the sentences have been examined. Exclusions (non-specific and non-referential readings) are found in Appendix 1.

*p-value: < 2.2e-16, 95% confidence interval: 90.8%-97.7% (using a binomial test)

5.1.4. Short and long form in the ACC-category for attributes

The biggest group among the ACC-tags is the ACC-GEN⁶⁷ (96), dealing with generally known referents. The greatest part of this chapter is therefore dedicated to this tag. The lesser used ACC-SIT (13; for situationally accessible referents) and ACC-INF (25; for referents that are accessible by inference from another referent) are treated in the final sub chapter.

Whereas ACC-SIT and ACC-INF display equal numbers of SF/LF, the ACC-GEN category initially displays a rather small number of SF, 9.4% (which will decrease to 5.4% after the analysis).

5.1.4.1. IS-tag ACC-GEN and long form attribute

There is not much of interest to say about the LF ACC-GEN group, other than that there is a certain discrepancy between article use in Greek and OCS LF here. LF is the expected form to find, as illustrated by (72).

- (72) и иде плъза къ **чръмъноуоумоу мороу**. да съконъчаютъ сѧ реченоу
стъимъ доухомъ. (CS, Paul the Simple, 58729)
 And he went, sailing to the **Black Sea** so that that which had been said by the **Holy Spirit** would come to pass.

The accessible referents lack articles in Greek more often than they do in the OLD category. In the LF ACC-GEN group that has Greek originals (70), 25.3% (22) correspond to anarthrous Greek constructions. The same percentage for OLD LF corresponding to anarthrous Greek is 4.8% (8), making this a significant difference⁶⁸. What is key to explaining the discrepancies in article/LF use among the ACC-GEN is that the amount of different referents is very small. Among those that display SF in OCS, we find four instances of the Holy Spirit (mentioned below). Among the anarthrous Greek ACC-GEN that correspond to OCS LF, we find the Eternal Life, *životъ večъnъ*, a total of 15 times, together with a couple of other well-known concepts. The OCS translators clearly thought that

⁶⁷ Recall that there is a certain amount of ACC-GEN “hidden” in the OLD category as well, since what is generally known is tagged as OLD when it is picked up again within the 13 sentence limit established in the corpus annotation. In other words, what we find here are first mention accessible referents.

⁶⁸ P-value < 0.0001, using a 2x2 contingency table and a Fisher’s exact test.

“eternal” should be marked by LF, even though Greek did not use an article. It is not uncommon for Greek to drop articles on very well-known referents (Napoli, 2009, p. 573), but OCS does not seem to follow this trend actively.

In general, among adjectives (since in this group only 10.3% are participles) modifying accessible heads, there is little variety among the lemmas. A fourth of the adjectives is made up by *večъnъ* “eternal”, 16.3% by *nebesъskъ* “heavenly”, and another 20.9% by *svetъ* “holy” and *člověčъskъ* “of-man/human” combined. Among the heads, we find that *životъ* “life”, *cesarъstvije* “kingdom”, *duchъ* “spirit”, *synъ* “son”, *bogъ* “god” and *ognъ* “fire” make up 67.4%, meaning that we are mainly dealing with the by now highly familiar referents Son of Man, Kingdom of Heaven, Holy Spirit, Eternal Life, Eternal Fire, etc.

The accessible category thereby shows a very similar LF percentage (94.6%) to that of the OLD referents (95.1%).

5.1.4.2. IS-tag ACC-GEN and short form attribute

ACC-GEN SF number a total of 9. As has been seen in the OLD-chapter, clearly known referents like God and the Holy Spirit *do* appear modified by SF, but it is rare. The Holy Spirit is responsible for four out of nine SF among the ACC-GEN. The fact that the Holy Spirit appears in SF when it is obviously a known referent could be explained in several different ways: 1) the Holy Spirit is such a well-established referent that marking it as definite in OCS was not considered necessary, cf. the English “God Almighty” which needs no article, and the fact that the Holy Spirit is always accompanied by the same adjective, *svętъ* “holy”, makes LF even more redundant as it acts more like a set expression, or 2) the Holy Spirit often appears in abbreviated form (as does God), saving both space and time for the scribe, where adding a LF to an abbreviated adjective seems somewhat pointless. Finally, 3) SF could be a result of the *type* of reference to the Holy Spirit, notably as “being filled by the substance of Holy Spirit”, which occurs twice in this group (and was also seen in chapter 5.1.3.2).

Among the remaining five SF there is one example (73) that illustrates some difficulties in discerning between accessible, generic and non-specific referents.

- (73) Подобъни сѣтъ отрочиштемъ **сѣдаштемъ** на тръжиштихъ. и приглашаѣштемъ
другъ друга. и глаголюштемъ. (CM, Luke 7:32, 40344)
homoi eisin paidiois **tois** en agora **kathēmenois** kai prosfōnousin allēlois, legontes
(20572)
They are like children **sitting** in the marketplace and calling out to each other.

The English version makes no reference to known children, i.e. the children that we just saw at the marketplace or the children that are usually found at the marketplace. The referent is *hypothetical* - children that *can* be found at a marketplace, i.e. a non-specific interpretation. This has, however,

been tagged ACC-GEN in Greek (probably due to the placement of the article, “children, **the** ones sitting...”). The SF in OCS indicates that the translator has probably chosen to view the children as having a non-specific referent. Another sentence is excluded on the same basis.

Among the ACC-GEN we also find the first mention of the “blazing furnace” mentioned in example (64) in chapter 5.1.3.2 (page 56), and the same argument used there is valid here: the translator saw this as an expression not involving a known furnace, but the concept of burning in a furnace in general. Both this ACC-GEN and the second mention OLD sentence are arthrous in Greek. The remaining two SF are the ones displayed below:

- (74) да събодеть сѧ реченое пѣркмь ꙗ҃лжштемь. (CM, Matt 13:35, 38820)
 hopōs plērōthē to rēthen dia tou profētou Hēsaïou **legontos** (15239)
 This was to fulfill what the prophet spoke:
 Literally: May be fulfilled that which was said by the prophet, **saying**:
- (75) онъ бѣ свѣтильникъ **гора** и свѣта. (CM, John 5:35, 51856)
 ekeinois ēn ho lukhnos **ho kaiomenos** kai fainōn (48647)
 John was a **burning** and shining lamp.

Example (74) is an example of a cataphoric participle delimiting the referent (that which was spoken by the prophet) by providing a direct quote (in Matt 15:35). For such uses we have previously seen LF, which we would expect here as well, if the explanation is not that the participle is an *adverbial* complement. The predicative⁶⁹ example (75) shows expected SF in *gorę* (*i svete*). The lamp has been tagged ACC-GEN because the Greek says “He was *the* burning lamp”, i.e. the lamp is referential, but OCS seemingly leaves the lamp non-referential. This cannot be considered erroneous.

In conclusion, one of the four sentences not referring to the Holy Spirit is an error. The other are rightly short and most likely belong in another IS-category, giving an overall SF percentage of 5.4%.

Table 17 LF/SF distribution in the ACC-GEN category for attributive adjectives and participles

| | SF | LF |
|--------------------------|----------|-------------|
| From corpus data | 9 (9.4%) | 87 (90.6%) |
| Dubious cases to exclude | 4 | 0 |
| Final analysis | 5 (5.4%) | 87 (94.6%)* |

Table displaying the original amount of SF/LF in the ACC-GEN category and the final percentage arrived at after all the sentences have been examined. Exclusions are referents that seem to be non-specific, found in Appendix 1.

*p-value: < 2.2e-16, 95% confidence interval: 87.8%-98.2% (using a binomial test)

⁶⁹ A small amount of predicative heads are found in the narrowed down dataset. These are instances where the head word is itself predicative, but has an information status of its own, i.e. it is referential and not just qualifying. There are 18 such instances altogether, with a majority (13) of OLD and ACC-GEN.

5.1.4.3. Strength in the ACC-INF and ACC-SIT categories for attributes

ACC-SIT (13) and ACC-INF (27) are small groups in the attributive subset, and will therefore only be commented upon briefly. In the ACC-SIT category there are both SF and LF (5 vs. 8). ACC-SIT is used in dialogue, when something is deemed to be identifiable by the hearer in the situation. This “something” naturally should be referential, something examples (76) and (77) can be argued *not* to be. Example (76) contains a deictic marker (“here”), but the place referred to is not necessarily referential; the adjective is qualifying. In (77), what the “good thing” consists in is self-explanatory based on the situation, and it does not need to be referential to get the meaning across, it is “good” as opposed to “bad”. This exact sentence was also seen as OLD in chapter 5.1.3.2, example (65) (page 57). As such, both the good thing and the empty place could be seen as non-specific and non-referential, and do not belong in the ACC-SIT category.

- (76) ...ѣко съде на **поустѣ** **мѣстѣ** есмѣ (CM, Luke 9:12, 40486)
hoti hōde en **erēmō** **topō** esmen (20721)
...for we are here in a **desolate** **place**.
- (77) **добро** бо **дѣло** съдѣла о мѣнѣ (CM, Mark 14:6, 37164)
kalon **ergon** ērgasato en emoi (7336)
She has done a **beautiful** **thing** to me.

There is, however, one difficulty in the SF ACC-SIT group, notably the presence of the interrogative pronoun *kakovъ*, which appears three times in the entire corpus. All of these are ACC-SIT, and show only SF in the variants *kakovo* and *kakova*. It is uncertain whether pronouns such as these behave as regular attributive adjectives when it comes to SF/LF. The demonstrative pronoun *takovъ* does show LF, but overall the number of occurrences is small. In the LF group of ACC-SIT, we find one instance of *takovychъ*. These pronouns/adjectives do not bestow a quality upon their head words (and do not individualize them) and neither *kakovъ*, nor *takovъ* is taken into account here.

Among the LF, we find typical situational use of ACC-SIT. Example (78) refers to a herd of pigs that has just been mentioned in the narrative (but not in the dialogue), i.e. that is present to the speaker and addressee, and where the LF indicates that both parties can identify said herd of pigs.

- (78) ...повели намѣ ити въ **стадо** **свиное** (CM, Matt 8:31, 50742)
...aposteilon hēmas eis **tēn agelēn** **tōn khoirōn** (47778)
...send us away into **the herd** **of pigs**.
- (79) отъпусти ѿ да шедъше въ **окръстїиныхъ** **селѣхъ** (CM, Mark 6:36, 50354)
apoluson autous, hina apelthontes eis **tous** **kuklō** agrous kai **kōmas** (6816)
Send them away to go into the **surrounding** countryside and **villages**.
(Greek uses an adverb and not an adjective here.)

There are two such examples, and the rest of the LF are strictly situational, as illustrated by example (79) where “the surrounding” only has meaning in reference to a center, which is the place of the utterance. LF dominates in the ACC-SIT category. There are instances of SF, but they are most likely

all non-specific. The very small number of examples makes it impossible to say if this is actually a significant majority.

The ACC-INF category shows nearly identical numbers of SF/LF (14/13). To illustrate what kind of sentences are found here, example (80) can be used. The “narrow door” in question refers to the entrance into the Kingdom of Heaven that has been previously discussed, and not to any concrete door. That it is therefore marked with LF is explained by the inference from what has previously been said regarding the Kingdom of Heaven and how difficult it is to enter it.

- (80) подвизаите са въннѣ сквозѣ **тѣснаа врата**. Ъко мѣноси ꙗко ѿвѣщѣхъ въннѣ и
 не възмогѣхъ. (CM, Luke 13:24, 51507)
 agōnizesthe eiselthein dia tēs stenēs thuras, hoti polloi, lego humin, zētēsousin
 eiselthein kai ouk iskhousin (21094)
 Strive to enter through the **narrow door**. For many, I tell you, will seek to enter and
 will not be able.

With referents such as the one above, which are identifiable inasmuch as they can be inferred from a referent that is already present, LF is to be expected. The ACC-INF tag in the corpus has been used in the annotation of Greek to explain *why* previously unmentioned NPs are marked as definite (forthcoming Haug et al., p. 13). For this reason, the cases of SF deserve more mention.

One feature that stands out in the SF group is the case distribution; we find 4 locatives and 3 instrumentals in this group (NA⁷⁰ is at 35.7%), whereas there are none among the LF, where 77% are made up of NA. This may, however, be purely incidental. The three instrumentals are all from similar sentences, speaking of how certain disciples eat “with unclean hands” (where said hands are inferred from said disciples), either in the dual or the plural. These all correspond to anarthrous Greek originals, and the focus here is not on the hands, but that they are unclean as opposed to clean, i.e. the interpretation is non-specific.

An example of *possibly* referential use of SF ACC-INF is found in (81), where the clothing is inferred from the two men, tagged NEW. However, most likely the focus here is on that the men are dressed *well*, not on the actual pieces of clothing. As such it should probably be NON-SPEC.

- (81) ...и се **мѣжа двѣ** стасте въ нихъ въ ризахъ бльштѣхъ сѧ. (CM, Luke 24:4,
 41532)
 ...kai idou **andres duo** epestēsan autois en esthēti astraptousē (48606)
 ...**two men** stood by them in **dazzling apparel**.

The ACC-SIT & ACC-INF categories contain both SF and LF. The SF seem to be mostly explained by the possibility of non-specific readings, in which case the ACC-tags could be argued to be erroneous

⁷⁰ Nominative and Accusative.

altogether. This is, however, a problem that could only be avoided by directly annotating the OCS corpus, instead of transferring the IS-tags from the Greek. Due to small numbers these results are taken as mere pointers and not presented in a table.

5.1.5. Short and long form in the NON-SPEC category for attributes

NON-SPEC is a category containing 148 adjectives and participles. SF is expected, but the raw data from the corpus gives a LF percentage of 20.8%. In the analysis, this percentage will be reduced to 10.9%.

As mentioned in chapter 4.2, in the IS-annotation of the corpus, non-specific NPs do not establish referents outside of the short-term embedding they are found in (forthcoming Haug et al., p. 3). Negated NPs and NPs cooccurring with conditionals and quantified expressions (tagged QUANT) are non-specific (Ibid., p 15). NON-SPEC⁷¹ is expected to show SF, and this sub-chapter will focus mainly on LF. Two typical non-specific uses showing SF are represented by the following sentences:

- (82) нѣсть бо дрѣво **добро** **творя** плода зѣла. Ни дрѣво **зѣло** **творя** плода добра.
(CM, Luke 6:43, 40283)
ou gar estin dendron **kalon poioun** karpon sapron, oude **palin** dendron sapron
poioun karpon kalon (20510)
For no **good tree** **bears** bad fruit, nor again does a **bad tree** **bear** good fruit.
- (83) И повелѣ а҃урилиянѣ принести дрѣво **велико**. (CS, Vita of Paul and Juliana, 81212)
And Aurilian commanded them to bring a **big tree**.

Example (82) is representative of a negated and hypothetical NON-SPEC referent. In (83) the tree has no specific referent until a specific tree has been brought forward and been introduced. Among the NON-SPEC there are a total of 115 adjectives and 32 participles. Only one of the participles is long, making the group of participles fit nicely with the hypothesis. However, this one exception deserves mention, since it shows use of LF in cooccurrence with *jedinъ*:

- (84) тако ꙗ҃лж вамъ. радость бѣвааѣ прѣдъ а҃нлы бжии. о единомъ грѣшъницѣ
кажштимъ сѧ (CM, Luke 15:10, 40926)
houtōs, legō humin, ginetai khara enōpion tōn angelōn tou theou epi heni hamartōlō
metanoounti (21180)
In the same way, I tell you, there is rejoicing in the presence of the angels of God
over one sinner **who repents**.
- (85) ꙗ҃лж вамъ ѣко тако радость бѣдетъ на ꙗ҃бсе. о единомъ грѣшъницѣ **кажшти** сѧ.
(CM, Luke 15:7, 40922)
I tell you, there will be more joy in heaven over one sinner who repents.

⁷¹ The NON-SPECIFIC category does not only contain “simple” non-specific expressions, but also the tags NON-SPEC-OLD and NON-SPEC-INF (inferable from a non-specific referent). Of these, there are 13 attributive: 3 SF OLD and 2 LF OLD; 3 SF INF and 5 LF INF. For these categories, LF could perhaps be expected, since it is referring back to something that is definite within the context of the utterance (but which has no referent outside of it), but based on this small amount of examples there seems to be no such trend. These are not included in the final statistics in table 18.

Jedinъ can have several different meanings: the *specific* “a certain”, on par with modern day use of the Russian *odin* (Čaburgaev, 1974, p. 247), or the pure numeral “one”. The latter is possibly the use in (84), if we are not to interpret it as “the one (specific) sinner who repents”. Seen this way, LF does not seem out of place, even though it is anarthrous in Greek. The same sentence has appeared already in Luke 15:7, with SF, as seen in (85). There is a possibility that the scribe saw the second mention of a sinner in Luke 15:10 as referring back to the first one in Luke 15:7, and therefore used LF. This will not be deemed an erroneous use of the LF⁷².

The adjectives, on the other hand, show some more variation: 30 (26.1%) show LF. Some of these can be excluded based on the fact that even though their heads may be non-specific, they themselves have the IS-tag ACC-GEN, and their heads are therefore NON-SPEC-anchored. There are 5 occurrences of *nebesъskъ*, “heavenly/of Heaven”. The referents in question are the “Birds of Heaven” (also discussed in chapter 5.1.1.1), where the birds may be non-specific but “Heaven” hardly is. Of all the 9 “Birds of Heaven” present in the corpus (corresponding to arthrous Greek constructions) there is no SF, and the 4 not tagged as NON-SPEC are tagged as KIND, which is perhaps a more accurate annotation, and they are therefore “moved” to this category. *večъnъ*, “eternal”, only appears in the LF in the corpus, which makes sense since all the referents are OLD or ACC-GEN, except the 6 found among the NON-SPEC, where it is questionable if they really belong. “Eternal Life” (4), “Eternal Death” (1) and “Eternal Judgment” (1) are not really non-specific referents since they refer to specific concepts. They would seem to belong in the ACC-GEN category. The “moved” group is found in Appendix 1.

Other things that trigger the use of LF are occurrences where “a holy spirit” has most likely been interpreted (by the OCS translator) as the generally known “the Holy Spirit” that we are more used to. The LF on a NON-SPEC *dchmъ stymъ*⁷³ indicates that the translator misunderstood the anarthrous Greek indefinite construction. These examples of “a holy spirit” can therefore not be put in the group of true non-specifics. The lemmas *večъnъ* and *nebesъskъ* as well as these two “holy spirits” (13 altogether) are excluded from the NON-SPEC set, resulting in 16.7% LF where concepts or set phrases and generic/non-specific vacillation is mostly to blame for LF.

⁷² This is a locative, and Lunt (2001, p. 157) claims that for participles, LF is more common in cases other than N or A. Based on corpus data for participles (where the amount of locatives is relatively small, 17 in total), LF reaches 81.26% for non-NA. In the NA-category, the percentage of LF is 74.2%, i.e. not radically different from non-NA. The difference is not statistically significant with a p-value of 0.4511 (using a 2x2 contingency table and a Fisher’s exact test). Participles in general seem to be inclined to appear in the LF. This may, however, be explained by what linguistic surroundings they typically appear in (generic, quantified), rather than just the fact that they have been formed from verbs.

⁷³ Sentences 51255 (token ID 585475) and 41626 (token ID 605229). See Appendix 1 for instructions on how to use these to access the sentences.

Examples of the difficulty related to *concepts* will be dealt with first. A notable feature among the remaining instances (17) is the occurrence of set expressions or concepts. Delimiting what a concept is poses some problems, but as a general rule it should be common occurrence and an NP that is not merely a random combination of a noun and an adjective. Many of the things referred to as “concepts” in this paper are represented by a single word in other languages, as for example “adultery”. A total of 10 sentences refer to “divorce”, “adultery”, “wedding clothes”, “unquenchable fire”, “arid places” (desert) and “millstone”. That these are marked with LF when we expect SF could indicate that their status is closer to that of something generally known, which could trigger LF regardless of IS. These are *not* excluded from the final statistics. As is seen in (86), the millstone is hypothetical, and certainly not definite.

- (86) ꙗже аще съблзнить единого отъ малыхъ сихъ вѣроужштиихъ въ ма. добрѣе
емоу естъ паче. аще обложатъ камень жръновъны о выи его. ꙗ вѣврѣгътъ ꙗ въ
море (CM, Mark 9:42, 36873)
kai hos an skandalisē hena tōn micron toutōn tōn pistevontōn, kalon estin auto
mallon ei perikeitai mulos onikos peri ton trakhēlon autou kai beblētai eis tēn
thalassan (7035)
Whoever causes one of these little ones who believe in me to sin it would be better
for him if a great millstone were hung around his neck and he were thrown into the
sea.

Two more possible concepts from Codex Suprasliensis also display LF, referring to valuable stones (gems), *kamenijemъ mnogocenъnyimъ* (87) and a common monastery, *obyštii manastyрь* (87). The common monastery can be explained in another way however: it could very well refer to a specific monastery in the area referred to as “the common monastery” (ACC-GEN), making the use of LF logical. Because of these various possible interpretations, (88) will not be considered erroneous.

- (87) ꙗ принесоша ꙗликоже бѣаше ѿ него коумирь. съребрѣны и златы. ѿукрашены
камениемъ многоцѣнънымъ (CS, Vita of Paul and Juliana, 81020)
And they brought what he had of idols, of silver and of gold, decorated with **valuable**
stones.
(88) аште хоштеши чрънецъ быти. вьлѣзи въ обыштии манастырь. иждеже ваште
братиа юсть. иже могутъ тебѣ и болѣзнемъ твоимъ послужити. (CS, Paul the
Simple, 58681)
Go, if you want to be a monk, and enter into a common monastery where there are
more brethren, who will be able to tend to you and your afflictions⁷⁴.

With 12 rather clear concepts (of which one may just be ACC-GEN) already discussed, 7 sentences remain to look at. For these, NON-SPEC/KIND variation prevails. They will be presented below.

- (89) егда же **нечисты дхъ** изидеть отъ члвка. прѣходить скозь бездѣнаа мѣста. (CM,
Matt 12:43, 38758)
hotan de to akatharton pneuma exelthē apo tou anthrōpou, dierkhetai di’ anudrōn

⁷⁴ Translation from: http://clover.slavic.pitt.edu:8080/exist/paul/data/paul_main.html (2013-02-28)

topō... (15177)

When the unclean spirit has gone out of a person, it passes through waterless places...

(89) shows an example of possible variation between non-specific and generic. This is a hypothetical statement, and is part of a section accounting for what happens when an unclean spirit exits its host's body. The referent is non-specific since it is found within an (implicit) modal embedding, but the sense is also generic; this is what unclean spirits do.

3 more of the unexpected LF are from CS. One of these examples is of the type that could be generic, referring to "good wives and daughters of men" (90), which could be "*the* good wives and daughters of men" (generic). A possible generic reading is less obvious for "respectable women" in (91). Only (90) is therefore "moved" to the KIND category.

- (90) бл҃дѣникъ сы ѡзлиха паче вѣсѣхъ чловѣкъ. живѣштиихъ. вида чловѣчьскы **добрыя** жены и дѣштери. тѣмъ похота ѡ въ лице бесѣдоуа ѡмъ (CS, Vita of Paul and Juliana, 81042)

Being a worse adulterer than all other living men he desired the **good** wives and daughters of men when he saw them and talked directly to them.

- (91) ...ѡже тѣмъ ѡзмѣненіѡмъ прѣлсти жены **доброродѣныя**. (CS, Vita of Paul and Juliana, 81444)

...who with that change (trickery) enticed **honourable** women.

All in all, only 2 more LF are removed from the NON-SPEC category on the basis that one is generic and the other very possibly referential, leaving us with 14 (10.9%) erroneous LF, of which 11 are possible concepts.

The conclusion that can be drawn from these data is that typical non-specific utterances should carry short attributive modifiers, but there are many instances of non-specific referents that possibly refer to something that is, in a manner, generally known, or where the adjective in question is formed from a unique entity (such as Heaven or the World), which provokes LF as it makes the NP "anchored" and more easily identifiable. It is difficult to give any definite percentages of LF in this category since several examples can be interpreted in two different ways, but approximately 10.9% is a good estimate, illustrating that SF dominates. For this number, "concepts" have been included. These are things that should, based on the semantics, show SF, but their possible status as "concepts" triggers LF. Generic cases have been excluded. The summarizing table is found on the following page.

Table 18 LF/SF distribution in the NON-SPEC category for attributive adjectives and participles

| | SF | LF |
|--------------------------|--------------|------------|
| From corpus data | 114 (79.2%) | 30 (20.8%) |
| Dubious cases to exclude | 0 | 16 |
| Final analysis | 114 (89.1%)* | 14 (10.9%) |

Table displaying the original amount of SF/LF in the NON-SPEC category and the final percentage arrived at after all the sentences have been examined. Exclusions include transfers to the KIND and ACC-GEN categories (found in Appendix 1).

*p-value: < 2.2e-16, 95% confidence interval: 81.5%-93.3% (using a binomial test)

5.1.6. Short and long form in the KIND category for attributes

According to Flier (1974, pp. 69, 160), who refers to generics as “generalized nouns” LF is expected on *singular* generic nouns in OCS. Rondestvedt (1986, p. 158) does not limit her statement to singulars only, but adds that the most common type of generic noun is a plural nominalized adjective or participle. In Ancient Greek, definite *plural* NPs can also be generic (forthcoming Haug et al., p. 12)⁷⁵. In the IS analysis, these nouns are given a specific tag, KIND, and should thereby be easy to pick up on. As was demonstrated in chapter 2.2.3, generic referents do not form a clear-cut category, and they pose certain problems in OCS as well. They are sometimes difficult to separate from NON-SPEC, and this should be kept in mind when speaking of “all” the generics in the corpus. Flier (1974, p. 160) points out that the cases of “generalized nouns” modified by attributive adjectives are not very numerous, and in the PROIEL corpus, there are merely 50, some of which belong in the NON-SPEC category (see Appendix 1).

A typical example of a generic LF can be found in the following sentence, with four singular generic referents (the good/evil person, good/evil treasure) and two plural (good, evil):

- (92) **Добры** члвкъ **отъ** **добраго** съкровишта **износить** **добраѣ** **и** **зълы** члвкъ **отъ** **зълааго** съкровишта **износить** **зълаа**. (CM, Matt 12:35, 38749)
ho agathos anthrōpos **ek** **tou agathou** thēsaurou **ekballei** **ta agatha**, **kai** **ho ponēros** anthrōpos **ek** **tou ponērou** thēsaurou **ekballei** **ponēra** (15167)
The good person **out of** **his** **good** treasure **brings forth** **good**, **and** **the evil** person **out of** **his** **evil** treasure **brings forth** **evil**.

In this group of words (a total of 50), there are 8 participles and 42 adjectives, with an almost equal distribution between sg/pl. The overall amount of SF (15, roughly a third) would seem to indicate that the separation between LF and SF is not crystal clear for generics. The final distribution however, *after* exclusions described below have been carried out, is very different:

⁷⁵ For more comments on PROIEL annotation, see chapter 4.2.

Table 19 Distribution of number and strength in the KIND category, after exclusions

| | SF | LF |
|-----------|----|----|
| SINGULARS | 3 | 17 |
| DUALS | 1 | 0 |
| PLURALS | 1 | 18 |

Of the referents found in table 19, 30 are arthrous in Greek and only half are singulars. Flier’s claim about only singulars being generalized does not seem to hold true. However, to arrive at the numbers above, some sentences need to be looked closer at. One group of words that are mostly tagged as generic (KIND) in the corpus are those that are accompanied by the adjective “such” (Greek *toioutos*, OCS *takovъ*), and here we find one such example in the SF. According to Flier’s data (1974, p. 168), *takovъ* is usually short, but does show some vacillation. In chapter 5.1.4.3 *takovъ* and *kakovъ* were excluded on the basis that they are not qualitative, regular adjectives, but of a pronominal nature (cf. *ъsjękъ*), and the one SF example is excluded here as well. The corpus only contains five occurrences of *takovъ* as an attribute (all short), and 3 nominalizations, of which 2 are long. For this particular type of referent it is difficult to argue for a logical use of LF.

Another point to be made regarding the group of SF is the frequency of the -ъsk suffix, which appears here 5 times in the adjective *člověčъskъ*, which, as has already been mentioned, behaves largely like a regular adjective, but does show some odd SF for both old referents and, as in this case, generics. Altogether, there are 15 cases of generic referents modified by -ъsk adjectives, the majority being “birds of Heaven” and displaying LF, but there are 5 SF as well. The example in (93) is interesting since the OCS translator has used the dual instead of the anarthrous plural in the Greek original. It is very improbable that “into (the) hands of men” was somehow interpreted as non-specific “into the hands of a man” by the translator. It is, however, a case of dual, and dual displays SF where it should have LF elsewhere as well⁷⁶. Furthermore, it immediately precedes the conjunction “and”, *i*, which is identical to the LF addition in the accusative feminine dual (-ѣi). Therefore, this may be a case of graphemic interference.

- (93) ѿко снъ члвскы прѣданъ бждеть въ ржцѣ члвчсцѣ. ꙗ оубыжтъ ꙗ ꙗ оубьенъ бывъ
(CM, Mark 9:31, 36856)
hoti ho huios tou anthrōpou para didotai eis kheiras anthrōpōn kai apoktenousin
auton (7017)
“The Son of Man is delivered into the hands of men and they shall kill Him...”

⁷⁶ See further comments on this in chapter 5.1.3.2.

The other SF -ьsk adjectives refer to “the commandments of men”, “the praise of men”, “mankind”, “the leaven of the Pharisees” and should all display LF.

For the remainder of the KIND SF it seems like the OCS translator has interpreted these sentences as non-specific based on the Greek, as in the examples below.

- (94) нъ вино **ново** въ мѣхы новы въливають. и обое съблюдетъ ся (CM, Matt 9:17, 38566)
alla ballousin oinon **neon** eis askous kainous, kai amfoteri suntērountai (14979)
Neither do men put **new wine** into old wineskins.
- (95) зане аште въ **сыръ дръвъ** си творятъ. въ соусъ что бждеть (CM, Luke 23:31, 41494)
hoi ei en tō **hugrō xulō** tauta poiouein, en tō xērō ti genētai; (21771)
For if they do these things in a **green tree**, what shall be done in the dry?

In (94), which is a sentence that appears twice, “old wineskins” has been annotated as NON-SPEC, but “new wine” as KIND. Both can just as well be non-specific, as seems probable from the lack of articles (see example (134) on page 88). For (95) it is similarly easy to see a non-specific interpretation, and the rest of the examples of SF for KIND tags display very similar sentences.

Among the three participles present in the SF group, all can be excluded on the basis of a possible non-specific interpretation. Altogether, 10 sentences are excluded with this motivation. In conclusion, LF for attributes to generic heads seems very plausible, even though the percentage is not as high as for other categories. It should be noted that *all* of the SF instances that remain after exclusions are due to the -ьsk suffix, and to the one lemma *člověčьskъ*. Compared to the percentage for ACC-GEN, the difference in SF/LF distribution is not significant with a p-value of 0.7851⁷⁷.

Table 20 LF/SF distribution in the KIND category for attributive adjectives and participles

| | SF | LF |
|--------------------------|------------|------------|
| From corpus data | 15 (30.0%) | 35 (70.0%) |
| Dubious cases to exclude | 10 | 0 |
| Final analysis | 5 (12.5%)* | 35 (87.5%) |

Table displaying the original amount of SF/LF in the KIND category and the final percentage arrived at after all the sentences have been examined. Exclusions in this category are cases that have most likely been interpreted as non-specific by the translator. The sentences can be found in Appendix 1.

*p-value: 1.383e-06, 95% confidence interval: 4,2% - 27,0% (using a binomial test)

5.1.7. Attributes to proper nouns

In the literature on OCS, there is general agreement that proper nouns get LF attributes. Proper nouns denote unique individuals and are only in exceptional cases indefinite, as in “I don’t know *a* Mary, but I do know *a* Maria” where the first is non-specific and the second specific. Just like with

⁷⁷ Using a 2x2 contingency table and a Fisher’s exact test.

other specific referents, we would expect in any sort of narrative to first be introduced to a persona when they, so to say, enter the stage, and then identify them based on the name they are given. Whereas with common nouns definite articles or, in the case of OCS, LF adjectives would signal that we are already familiar with said referent, the proper noun does that all on its own. Some new referents may not even need an introduction, if said persona is someone who is very well-known in the cultural context, which is the case with certain biblical characters. Adjectives and participles should then logically show a very strong tendency to be long when accompanying proper nouns, i.e. they should agree with the inherent definiteness of the proper noun. This is also the case, as is demonstrated by the data from the corpus. For more details on appositive adjectives and participles to proper nouns, of which there are far more, see chapter 5.3.1.

An examination of all attributes that modify a proper noun gives the following results: 3 SF and 20 LF⁷⁸. Two of the SF are easily explained, since they are obviously indefinite, albeit including a proper noun, giving a SF/LF distribution of 1/20, i.e. a very strong 95.2% LF.

- (96) въстанѣтъ бо **лъжи хръсти**. и лъжи пророци. и дадѣтъ знамениѣ и чюдеса. да прѣльстѣтъ аще възможѣно избѣраныѣ. (CM, Mark 13:22, 37136)
 egerthēsontai de phseudokhristoi kai pseudoprofētai kai poiēsousin sēmeia kai terata pros to apoplanan, ei dunaton, tous eklektous (7305)
 For **false Christs** and false prophets will appear and perform signs and miracles to deceive the elect--if that were possible.

The Christs referred to are obviously not *the* Christ, and have been tagged as non-specific. The two sentences are duplicates from different texts. Another sentence, (97), denoting an accessible referent makes use of a -ъsk suffix formed from the name of a place, Tiberias (rendering the adjective *tiveriĕdъskъ*). As was mentioned in chapter 5.1.1.3, this group of -ъsk adjectives show much more erratic behavior than the rest of them, with many instances of SF where LF is expected, as in (97):

- (97) по сихъ иде ѿсь на онъ полъ морѣ галилъѣ (proper noun) **таверѣдъскы** (CM, John 6:1, 41914)
 meta tauta apēlthen ho Iēsous peran tēs thalassēs tēs Galilaias **tēs Tiberiados** (22211)
 After these things Jesus went across the Sea of Galilee which is the Sea of Tiberias.

The amount of proper nouns with attributive adjectives is not very great and the group only contains adjectives (all participles are appositions), but the numbers confirm the hypothesis that known elements get LF attributes.

⁷⁸ There *are* more attributes on proper nouns, but several of these happen to carry the suffixes that only display short form, and have been excluded automatically.

5.1.8. Word order for attributes

Dimitrova-Vulchanova & Vulchanov (2009) suggest that word order plays a role in definiteness for OCS, based on the texts they have examined. They claim that LF adjectives are expected to appear post-nominally. For the NT texts available in the PROIEL corpus, word order almost completely follows the Greek originals, and both SF and LF adjectives and participles appear largely post-nominally.

Table 21 Word order in Greek and OCS

| | | OCS Pre-nom. | Greek Pre-nom. | OCS Post-nom. | Greek Post-nom. | NGO* Pre-nom. | Post-nom. |
|---------|----|-----------------|-------------------|------------------|--------------------|------------------|-----------|
| | | | | | | | |
| NEW | SF | 11 | 11 | 52 | 52 | 1 | 9 |
| | LF | 4 | 4 | 32 | 32 | 5 | 4 |
| OLD | SF | 5 | 4 | 13 | 14 | 1 | 2 |
| | LF | 25 | 25 | 132 | 132 | 9 | 8 |
| ACC-GEN | SF | 0 | 0 | 9 | 9 | 0 | 0 |
| | LF | 11 | 11 | 59 | 59 | 6 | 11 |

*No Greek original

As becomes obvious from the table, OCS and Greek word order is more or less identic. There *is* a small number of mismatches that do not show in the table, amounting to a total of 3 mismatches in the NEW category, 3 among the ACC-GEN and 1 among the OLD. These are the following:

PRE-NOMINAL IN OCS, POST-NOMINAL IN GREEK:

na **dobrě** zemi vs. epi tēn gēn **tēn kalēn** (OLD⁷⁹ short form)

na **sty** dchъ vs. to pneuma **to hagion** (ACC-GEN long form)

nečistomъ dchmъ vs. en pneumatī **akathartō** (NEW short form, appears twice)

POST-NOMINAL IN OCS, PRE-NOMINAL IN GREEK:

na dchъ **sty** vs. eis **hagion** pneuma (ACC-GEN long form)

telesъ **pitomy** vs. ton **sitouton** moskhon (ACC-GEN long form)

na treve **zelene** vs. tōn **hlorō** hortō (NEW short form)

There are as many mismatches in each direction, which is why they do not show in the table, making it seem random rather than like any sort of trend for OCS to deviate from Greek word order.

Since Codex Suprasliensis is a younger text, some comments on word order patterns there is also of interest. CS shows a slight increase in pre-nominal attributes, with 33.3%⁸⁰, as opposed to the

⁷⁹ This very expression has been argued to be NON-SPEC in this thesis, not OLD.

⁸⁰ P-value = 0.002985, 95% confidence interval: 23.5% - 44.5% (using a binomial test).

21.4%⁸¹ found in CM. Of the pre-nominal attributes in CS, 63.6% show LF where SF is expected, i.e. for NEW and NON-SPEC referents, but there are so few NPs (11) that these results are statistically insignificant⁸². The percentage for unexpected pre-nominal LF in CM is at 25.6%⁸³ (10).

When more texts from CS are added to the corpus, something more conclusive can be said about any changes in word order. That the pre-nominal position should primarily be occupied by indefinite modifiers, as indicated by Dimitrova-Vulchanova and Vulchanov, does not seem true. That word order should help mark definiteness in Early Slavic is also refuted by Rondestvedt (1986, p. 225).

5.1.9. Conclusion regarding attributes

Attributes appear to largely behave as expected by the IS-tags, i.e. SF for NEW and NON-SPEC, and LF for OLD, ACC and KIND. Once cases that can be read in multiple ways have been excluded, the expected form shows very high percentages of over 90% in almost all categories. The sorting carried out in chapter 5.1 brings to light certain difficulties related to transferring IS-annotation from the Greek originals, especially in the non-specific and generic gray zone. There is no overwhelming sign of plurals behaving more erratically than singulars, and overall, the amount of errors is rather small, *too small* to say anything significant about a sg/pl division. In this chapter it has also been found that familiar referents are not very frequently marked by SF modifiers. Word order has been shown to follow Greek very closely, indicating that most likely no sort of definiteness coding is found there. In the section on generic referents we saw that there were both singulars and plurals, which goes against Flier's claim. As for denominal adjectives derived from proper nouns, we have only looked at those formed with the -ьsk suffix, among which most show SF/LF as expected. Adjectives derived from peoples and places (partly derived from proper nouns) are an exception.

5.2. Nominalizations

Nominalizations are claimed to show more use of LF than what the semantics/pragmatics would have them do, especially in the plural, according to Flier (1974). In this chapter, these theories are checked against the data from the corpus. After various categories have been examined, there is no statistically significant⁸⁴ unmotivated increase in LF use for nominalizations, nor does number seem to have any great effect on the use of LF rather than SF. Rather, most uses of LF in categories where we would expect SF can be easily explained. For those categories where we expect to find only LF, the nominalizations show very few SF. The table below (with unanalyzed data) summarizes the

⁸¹ P-value < 2.2e-16, 95% confidence interval: 18.5% - 24.5% (using a binomial test).

⁸² P-value = 0.5488 (using a binomial test).

⁸³ P-value = 0.003378 (using a binomial test).

⁸⁴ See the summarizing table in chapter 6 for precise statistical results comparing the SF/LF distribution for attributes and nominalizations.

occurrences of SF/LF and the distribution of adjectives/participles vs. number for nominalizations. The purpose of this table is primarily to show the distribution of sg/pl and to avoid unnecessary detail in the subsequent chapters.

Table 22 Summary of the distribution of number for adjectival and participial nominalizations

| | SF | Adjectives | | | Participles | | | LF | Adjectives | | | Participles | | |
|----------|----|------------|----|----|-------------|----|----|----|------------|----|----|-------------|----|----|
| | | SG | DU | PL | SG | DU | PL | | SG | DU | PL | SG | DU | PL |
| NEW | | 6 | 1 | 3 | 2 | 0 | 1 | | 3 | 0 | 3 | 14 | 0 | 7 |
| NEW-a* | | 6 | 0 | 1 | 5 | 0 | 4 | | 13 | 0 | 6 | 9 | 0 | 10 |
| OLD | | 9 | 0 | 5 | 3 | 0 | 1 | | 21 | 5 | 19 | 48 | 3 | 29 |
| ACC | | 1 | 0 | 3 | 3 | 0 | 1 | | 14 | 0 | 25 | 18 | 0 | 30 |
| NON-SPEC | | 30 | 0 | 8 | 6 | 0 | 0 | | 7 | 0 | 31 | 5 | 0 | 8 |
| KIND | | 6 | 0 | 6 | 0 | 0 | 1 | | 11 | 0 | 48 | 1 | 0 | 7 |
| QUANT | | 3 | 0 | 6 | 5 | 0 | 2 | | 2 | 0 | 8 | 84 | 0 | 60 |

*NEW-anchored

5.2.1. Short and long form in the NEW category for nominalizations

This is a rather small group containing 40 nominalizations. From all that has been said so far, we would continue to expect NEW referents to be marked with SF, headless or not. According to the literature, OCS shows a trend towards using LF for nominalizations regardless of whether the referent in question is identifiable or not. The large amount of LF in the NEW category, 67.5% (27), seems to back this claim up, but as will be seen, there are usually good explanations for the use of LF. After all exclusions have been made, there are not many nominalizations left as truly NEW, and the small numbers leads to a 20% (3) percentage of LF.

5.2.1.1. IS-tag NEW and short form nominalization

There are a mere 12 instances of NEW nominalized referents that show SF in this group. This, which we would expect to be the “norm” for previously unknown referents, is represented by (98):

- (98) Тѣгда привѣса къ немуу **бѣсънжѣщѣ** са слѣпъ и нѣмъ... (CM, Matt 12:22, 38730)
 tote prosēnekhthē auto **daimonizomenos** tuflos kai kōthos (15145)
 Then a **demon-oppressed man** who was blind and mute was brought to him...

In (98) there is a SF nominalization introducing a new referent, and the nominalization is also modified by two SF attributes, *slěpъ* and *němъ*. OCS seems to follow the Greek article use closely in this SF group, since all except one are anarthrous. The arthrous example in (99) is somewhat peculiar, as it corresponds to an expression of the type “it’s strange/wonderful”. What the speaker is referring to is the fact that even though Jesus accomplishes wonders, little is known about him.

- (99) о семь бо **дивъно** естъ. ѿко (CM, John 9:30, 42298)
 en toutō gar **to thaumaston** estin, hoti... (22625)
 Det var **underligt**, svarade mannen.
 C'est **étonnant**, répliqua l'homme.
 Why, this is **an amazing thing**!

As can be seen, this is not necessarily a referring expression, and *divъno* is most likely predicative. It could also be the subject, and still show a justifiable SF. The definite Greek original has the cataphoric meaning of “For in this lies *the* amazing thing, **that...**”, but the OCS translator does not seem to have chosen to translate it this way, but rather as “for in this it is amazing, that...”.

5.2.1.2. IS-tag NEW and long form nominalization

In the LF-group, numbering a total of 27, we find many nominalizations for which there is nothing strange at all about their LF use. These nominalizations (a total of 14 participles) correspond to restrictive relative clauses of the sort “he who said”, “those who did”, etc., that is to say a referent who is strictly speaking new in the narrative, but who is further individualized by the very nominalization denoting it or further information immediately provided. A notable example in this group is *rečenoje* in (100), which appears a total of 6 times. All of the examples of *rečenoje* are delimited (or anchored) by “the prophet” or “the Holy Spirit”, and are furthermore cataphoric. That these do not instead appear in the anchored group (discussed in chapter 5.2.1.3) is explained by how data is extracted from the corpus, and was mentioned in chapter 4.2. Example (101) also shows a nominalization that is further delimited by a dependent, notably “the wedding”, which is OLD. 10 out of the 14 have such dependents that anchor them.

- (100) да събждеть сѧ **реченое** ꙗкоже ꙗсаиѣ ꙗвлѣхъ (CM, Matt 12:17, 38724)
 hina plērōthē **to rēthen** dia Hēsaïou tou profētou legontos (15139)
 This was to fulfill **what was spoken** by the prophet Isaiah:
- (101) ...и посылѧ рабы свои призвати **зѣванныи на бракъ**. CM, Matt 22:3, 39337)
 ...kai apesteilen tous doulous autou kalesai **tous keklēmenous eis tous gamous**
 (15756)
 ...and sent his servants to call **those who were invited to the wedding feast**.
- (102) ꙗ **мимо ходаште** хоулѧхъ и. ꙗ поклѧвахъте главами своими. ꙗ ꙗлше (CM, Mark 15:29, 37316)
 kai **hoi paraporeuomenoi** eblastēmoun auton kinountes tas kefalas autōn kai legontes (7493)
 And **those who passed by** derided him, wagging their heads and saying... (literally: near, right by)

The other four contain one example of an elliptic anchoring PRO-DROP object in (102), i.e. “those who passed by *him* derided him”, one with a cataphoric marker (*pisanoje*, “that which is written”)

and two without any particular dependents (“the builders”, referred to in a quote, and “the things that are to come”⁸⁵). None of these 14 actually shows unexpected LF.

Three other sentences are variants of “on the right side”, i.e. *o desnqjɔ*. This belongs in the category of absolutes, where “the right” is identifiable in relation to “the left”, and is always relative to something. Finding LF here is therefore not surprising, and there is only a single instance among dozens of this expression showing SF. Furthermore, if “on the right side” is really to be considered a new referent is questionable. Most likely, these belong in the ACC-INF category since they can usually be identified from the context, and they are excluded based on this.

Among the rest of the LF (17), there are but a few examples where LF is unexpected or odd. For example (103) (which is repeated twice), it is questionable whether NEW is really the appropriate tag; it is clearly anchored by the presence of the possessive “of the kingdom of God”. Anything other than LF would appear strange. It is not a question of “a secret” among many, which is clear from the wider context in Mark 4.

- (103) вамъ естъ дано вѣдѣти **тайнаа** цѣрствиѣ бжиѣ. (CM, Mark 4:11, 50279)
 humin **to mustērion** dedotai tēs Basileias tou Theou (6667)
 To you has been given **the secret** of the kingdom of God
- (104) ...и остависте **тажъшаа** закона сѣдѣ и милость и вѣрж (CM, Matt 23:23, 51046)
 kai afēkate **ta barutera** tou nomou, tēn krisin kai to eleos kai tēn pistin (15852)
 ...and have neglected **the weightier matters** of the law: justice and mercy and
 faithfulness.

In (104) there is a comparative. The matters at hand are “weightier” in relation to some less important matters and are an identifiable part of a whole, and therefore counted as definite. In addition, the matters are anchored by the presence of the generally accessible “law”, as well as by a list of said matters, which is provided straight away.

- (105) Се изиде **сѣи** сѣять. (CM, Mark 4:3, 36510)
 idou exēlthen **ho speirōn** speirai (6657)
 Behold, **a sower** went out to sow.
- (106) гласъ **въпиѣщааго** въ поустыни. (CM, Mark 1:3, 36361)
 fonē **boōntos** en tē erēmō (6502)
 the voice of **one crying** in the wilderness

Example (105), which is part of a parable, appears three times. All of these examples are also arthrous in Greek. This being a parable is probably what explains the use of a seemingly definite referent for something that is textually new, and the article/the LF is thereby used as a narrative device (*in medias res*). The “one crying in the wilderness” in (106) is then the only example – which also appears three times – which shows truly unexpected LF. This sentence corresponds to an

⁸⁵ Sentence and token IDs: John 16:13, 42721 (617834) & Mark 12:10, 37038 (547426).

anarthrous Greek construction, so LF here is not merely a case of copying the Greek, which makes it all the more curious that it appears in LF three times. The OCS translator must have interpreted it as “he who cries in the wilderness/the one crying in the wilderness” (the “wilderness” furthermore being an anchoring factor). Even though such an interpretation is not *necessarily* erroneous, these three will count as “LF where we expect SF” in the table at the end of the chapter.

Number seems to play very little role here, and the only LF occurrences we have left are singular (and 3 duplicates of the same sentence). In the final percentages, cases of participles corresponding to restrictive relative clauses, of which many are anchored, have not been included since these cannot be considered erroneous. What becomes most apparent from this radical narrowing down in the LF category is that the NEW tag does not always denote unidentifiable referents, which is what we end up with in the “Final analysis”-row in table 20. It also does not in any way seem like nominalizations display random use of LF when there should be SF. The number of occurrences is so small that the results are barely significant (see the p-value in table 23).

Table 23 LF/SF distribution in the NEW category for nominalizations

| | SF | LF |
|--------------------------|-------------|------------|
| From corpus data | 13 (32.5%) | 27 (67.5%) |
| Dubious cases to exclude | 1 | 24 |
| Final analysis | 12 (80.0%)* | 3 (20.0%) |

Table displaying the original amount of SF/LF in the NEW category and the final percentage arrived at after all the sentences have been examined. Exclusions in this category feature (often anchored) restrictive relative clauses, one instance of a referent that can be quantified and one that can be generic (found in Appendix 1).

*p-value: 0.03516, 95% confidence interval: 51.9%-95.7% (using a binomial test)

5.2.1.3. Short and long form for NEW-anchored referents

The NEW-group for nominalizations contains a large amount of anchored referents (54), which get the tag NEW-anchored. This group *should* be even larger, since several anchored referents that have not been picked up as such in the extraction of the dataset were encountered in the previous sub-chapter. Anchored referents initially display 75.9% LF, and after exclusions, 93.2%, which is not unexpected since these nominalizations represent things that are more easily identifiable than purely new referents. For attributes (which show a total of 72 anchored NEW referents), the final LF percentage for anchored referents was also 93.3%. A typical anchored referent is found below in (107), where the anchoring factor is the *jb*, “him”.

- (107) възърѣвъ же великѣи антонии на юношѣ. глагола **ВОДАШТИИМЪ** ѿ. (CS, Paul the Simple, 58695)
The great Anthony glanced at the young man, and said to **those who were leading him...**

The 13 SF anchored referents show 11 singulars. Out of these 13 anchored referents, we find that 9 can be explained, and only 4 should really display LF. There are three instances of NEW-tags that should be ACC-INF (see Appendix 1), two comparisons where something clearly indefinite is compared to a referent tagged as OLD (meaning these are not really anchored), one specific indefinite cooccurring with *jedinъ*, and two referents that are erroneously picked up as anchored. These examples can be found in Appendix 2. Finally, in (108), what anchors the “dead man” is the “his” in “his mother”, which is tagged OLD and inferred from the dead man himself. This is not a proper anchoring.

- (108) І се изношаахъ **оумерошъ** (NEW) ѿнъ иночадъ матери своеи (OLD) (CM, Luke 7:12, 40311)
 kai idou exekomizeto **tethnēkōs** monogenēs huios tē mētri autou (20538)
 Behold, **a man who had died** was being carried out, the only son of his mother.

All of the anchored referents that are excluded on the basis that they are not truly anchored are anarthrous in Greek. An interesting fact about the four remaining – and truly anchored referents – that should show LF, is that two of them feature cases of SF plus an “i” directly following it, as in example (109). Here the *јъ* has been interpreted as “him” in the annotation, but could possibly be the LF ending, or just a reason why the scribe missed to place an extra *јъ* after the participle. Due to the ambiguity here this example is counted as a LF. The anchoring factor here is a pro-drop object tagged OLD in the Greek original. An obvious erroneous SF is found in example (110), which also features an OLD pro-drop oblique (in the OCS translation this would be a complement to the adverb *poslědъ*).

- (109) І **видѣвъ** и съвѣдѣтельствова (CM, John 19:35, 42916)
 kai **ho heōrakōs** memarturēken (23299)
 He who saw it has borne witness.
 (110) І послѣдъ **градъште** боѡахъ сѧ (CM, Mark 10:32, 36932)
 hoi de akolouthountes efobounto (7095)
 And **those who followed** (him) were afraid.

As such, LF is quite obviously what is to be expected for anchored nominalizations. See the table on the following page.

Table 24 LF/SF distribution in the NEW-anchored category for nominalizations

| | SF | LF |
|--------------------------|------------|---------------------------|
| From corpus data | 13 (24.1%) | 41 (75.9%) |
| Dubious cases to exclude | 10 | 1 |
| Final analysis | 3 (6.8%) | 41 ⁸⁶ (93.2%)* |

Table displaying the original amount of SF/LF in the NEW category and the final percentage arrived at after all the sentences have been examined. Exclusions in this category feature cases where formal anchoring in no way affects the identifiability of the referent, erroneous IS-tags and referents that are not anchored.

*p-value = 1.618e-09, 95% confidence interval: 81.3%-98.6% (using a binomial test)

5.2.2. Short and long form in the OLD category for nominalizations

Nominalizations in the OLD category are quite numerous at 140, with a clear majority of LF (89.3%).

The LF percentage will increase further to 93.9% with the exclusions carried below.

5.2.2.1. IS-tag OLD and long form nominalization

Since LF must be supposed to be the norm for OLD nominalizations, some general comments will be made here. Sentence (111) illustrates the typical OLD LF nominalization, of which there are 125, where “the holy ones” is anaphoric and refers back to a previous mention of Juliana and Paul:

- (111) мнози же отъ съчець мѣчаштиихъ **сѣя** повелѣниѣмъ аурилиянемъ. видаште
силы бываемыа богомъ. вѣроваша къ богу (CS, Vita of Paul and Juliana, 81007)
Many of the executioners, who were torturing **the holy ones** at the orders of Aurilian
believed in God seeing the strength that was with God.

13 of these LF correspond to anarthrous Greek constructions, and these will be looked at more closely here. A repeated construction that appears four times in this group is “from the dead” and another two “of the dead”, where the OCS translator either saw this as anaphoric (picking up previous mentions of the same “dead”), or as a generic reference. In the Greek annotation, these mentions of “dead” refer back to other – anarthrous – mentions of “dead”, and it is mainly a question of “rising from the dead”. As per the PROIEL annotation guidelines (forthcoming Haug et al., p. 7), generic referents that are picked up are tagged as OLD, but they *are* still generic, and OCS shows LF as expected.

Two other sentences feature “the highest” (here: God), which in Greek is expressed by *hypsistos*⁸⁷, and in OCS always with the LF *vyšbnii* (see comments on comparatives/superlatives in chapter 3.5.2). Another two feature the already discussed *desnojq* (“right”) and *šjujqjq* (“left”), and one features a

⁸⁶ That the final analysis contains 41 LF is not a mistake: one of the nominalizations found in the SF-group is most likely a LF; it is removed from the SF group and added into the LF group, while another LF has been removed on the grounds that it belongs in the ACC-INF category.

⁸⁷ Cf. the comments in chapter 5.1.4.1 on how Greek drops articles on well-known referents whereas OCS typically does not.

coordination where the Greek only marked one coordinated part with an article. There is one more LF corresponding to an anarthrous Greek original, where staying close to the Greek would dictate SF:

- (112) кто же есть съ о немъже азъ слышж **таковаа**. (CM, Luke 9:9, 40479)
tis de estin houtos peri hou akouō **toiauta**; (20714)
Who is this about whom I hear **such things**?

The LF in (112) is somewhat odd, unless the interpretation is “these things”. The Greek *toiauta* here refers back to things that “were happening”. As such, it is inferable from the situation what these things are, and this should possibly have been tagged ACC-INF – for which we would expect LF. Usually, *toiauta/takovъ*⁸⁸ is tagged KIND as it induces a kind-referring interpretation. However, here we must assume that it refers back to specific “things”.

Among the nominalizations marked as OLD-INACT, 12 (out of 13) are long. The total percentage of LF for OLD nominalizations is at 92.7%, as displayed in the table at the end of the next section.

5.2.2.2. IS-tag OLD and short form nominalization

The *original* share of SF in the OLD category at 10.7% is very similar to that for attributes, which was 10.4%. The 15 SF in this group correspond to 10 arthrous and 5 anarthrous Greek constructions, of which 9 will be considered erroneous in the final analysis. We will first examine the anarthrous ones, where there is a greater possibility of the OCS translator just having followed the article use in Greek, and then the SF that correspond to Greek examples with definite articles.

The anarthrous group is very small: the five occurrences are actually represented by two unique sentences: one that is repeated three times in the codices, and another containing two SF, and which happens to be the only instance of “on the right/left” showing SF in the entire corpus.

- (113) а еже състи о **деснж** и о **шжж** мене нѣсть мнѣ сего дати. (CM, Matt 20:23, 50953)
to de kathisai ek **dexiōn** mou kai ex **euōnumōn**, ouk estin emon touto dounai (47909)
But to sit **at my right hand** and **at my left** is not mine to grant.
- (114) добры рабе и благы ѿ вѣрне о **малъ** бѣ вѣренъ. (CM, Matt 25:21, 51099)
eu, doule agathe kai piste, epi **oliga** ēs pistos (47973)
Well done, good and faithful servant. You have been faithful over **a little**.

The SF in (113) is a rarity, as 19 out of 20 occurrences of this expression display LF. The sentence presented in (114) is also found in Matt 21:23 and Luke 19:17. “A little” does, indeed, have a referent further back in the text (a “mina” and two and five “talents” respectively, that is coins), but just as with “the blazing furnace” in chapter 5.1.3.2, example (64) (page 56), this is most likely a case of emphasis and the expression is non-specific or non-referential. The ides is not to refer back to the already mentioned coins, but to make a point out of the good servant having been faithful in “a little”

⁸⁸ For comments on *takovъ/kakovъ* as attributes, see chapters 5.1.4.3 and 5.1.6.

as opposed to “a lot”, which is also the nuance the English translation carries. As such, (113) is considered erroneous, but (114) and its duplicates are not, and are excluded from the final statistics.

Among the arthrous SF (10) there are a total of 7 that can be outright deemed erroneous, having clear referents further back in the text and with no clear “attenuating circumstances” as to why they display SF. One is found below in (115), where “he who comes” refers back to “the son of David”. The rest can be found in Appendix 2. These, together with the two nominalizations found in example (113) are what make up the final analysis-LF share in table 25.

- (115) Сѡсанна ѿнови давыдовоу. бл҃гослвнѣ **градѣ** въ імаѣне. ѡсанна въ вѣшнѣиихъ.
 (CM, Matt 21:9, 39255)
 ōsanna tō huiō Daveid, eulogēmenos **ho erxomenos** en onomati kuriou, ōsanna en
 tois hupsistois (15672)
 “Hosanna to the Son of David! Blessed is **he who comes** in the name of the Lord!
 Hosanna in the highest!”

The remaining 3, presented below, show some curiosities. Two contain the expression “fear of the Jews” (and does have antecedents, but this could be irrelevant), which could, possibly, have been interpreted as indefinite and as simply “a (general) fear of Jews”, where the nominalized possessive *ijudeiskъ* represents “the Jews”. As was mentioned in chapter 5.1.1.3, the -ъsk suffix on adjectives denoting peoples and places also behaves erratically, and especially the *ijudeiskъ* adjective. Therefore, it is difficult to know whether the SF is there for a pragmatic reason or just as a result of the adjective very rarely displaying LF.

- (116) никотерѣи же оубо ѣвъ глааше о немъ. страха ради **июдеиска** (CM, John 7:13,
 42042)
 oudeis mentoi parrēsia elalei peri autou dia don fobon **tōn loudaiōn** (22343)
 Yet for fear of **the Jews** no one spoke openly of him.
- (117) азъ же не искж славы моеѣ. естъ **ишта** и сѣда (CM, John 8:50, 42217)
 ego de ou zētō tēn doxan mou estin **ho zētōn** kai krinōn (22538)
 Non, je ne recherche pas la gloire pour moi-même: c'est un autre **qui s'en préoccupe**
 et il me rendra justice.
 Yet I do not seek my own glory; there is One **who seeks it**, and he is the judge.

(116) is not considered an error since it can be non-referential, and is excluded from the unexpected SF. In (117), the OCS translator seems to have opted for a solution of the type used in English and French, with a new and seemingly unknown referent introduced (a “someone”, “un autre”), even though this referent is obviously God. The Greek has a definite “the one who seeks and judges”.

As was mentioned in chapter 5.2.2.1, there is one instance of an OLD-INACT tag with SF among the nominalizations, presented below in (118). That *sucho*, i.e. “land” should be marked as OLD to begin with is perhaps not the best choice, but even as ACC-GEN or ACC-SIT, LF on a nominalization would be expected. There are no other occurrences of *sucho* meaning “land” in the corpus, so it is difficult

to say that this is a set phrase of the adverbial sort only occurring in SF. It will be considered erroneous here, even though it could be read as non-referential, i.e. land as opposed to water.

- (118) ι извезъше корабль на **coyxо** и оставъше все въ слѣдъ его идѣ (CM, Luke 5:11, 40173)
 kai katagagontes ta ploia epi **tēn gēn**, afentes panta ēkolouthēsan autō (20396)
 And when they had brought their boats to **land**, they left everything and followed him.

The final percentage of SF in the OLD category ends up being 6.7%, and thereby somewhat greater than for attributes, for which this percentage was at 4.9%. If the example concerning *sucho* is also excluded, the percentage ends up at 6.1%. This is not a statistically significant difference⁸⁹.

Table 25 SF/LF distribution in the OLD category for nominalizations

| | SF | LF |
|--------------------------|------------|--------------|
| From corpus data | 15 (10.7%) | 125 (89.3%) |
| Dubious cases to exclude | 6 | 0 |
| Final analysis | 9 (6.7%) | 125 (93.3%)* |

Table displaying the original amount of SF/LF in the OLD category and the final percentage arrived at after all the sentences have been examined.

*p-value < 2.2e-16, 95% confidence interval: 87.6%-96.9% (using a binomial test)

5.2.3. Short and long form in the ACC category for nominalizations

The ACC-category shows remarkable uniformity, with LF at 92.6% for all three categories (a total of 95 nominalizations) before any exclusions. Since the ACC-GEN category has so very few SF, there will be no separate sub-chapters for SF and LF. Nominalized ACC-INF, on the other hand, are very numerous, and not put together with the ACC-SIT. The table at the end of the chapter displays the percentages for all the ACC categories combined, which will show that LF reaches 95.6%.

5.2.3.1. Strength in the ACC-GEN category for nominalizations

Among the 39 ACC-GEN nominalizations, there are but 2 SF. A typical LF use is represented by (119).

- (119) дѣхъ сѣи вѣидеть въ та и сила **вышѣняго** оубѣснить та. (CS, Vita of Paul and Juliana, 81148)
 The Holy Spirit will go into you and the power **of the Highest** will bless you.

There are four LF ACC-GEN that correspond to anarthrous Greek originals. These include three instances of the superlative “the highest” in the genitive sg/pl, *vyšneego/vyšniichъ*, and one “those born of women”, where the Greek anarthrous *en gennētois* is rendered by the LF *roždenychъ* in OCS. That the Greek has no article changes very little, since the indefinite “people born of women” and

⁸⁹ P-value = 0.9349, using a 2x2 contingency table and a Fisher’s exact test.

the definite “those born of women” both have the same referent: everyone. Superlatives (which in OCS are formally comparatives), as we have seen in chapter 3.5.2, show a strong tendency towards LF. All these OCS LF are therefore to be expected, anarthrous Greek originals or not.

Among the two SF we find “fear of the Jews”, an expression also encountered among the SF OLD nominalizations. The same arguments hold here: this could have a non-specific interpretation, or the -ъsk suffix (and/or this particular lemma) could explain the use of SF. The second SF is the following:

- (120) се иже въ одежди славѣнѣ ꙗ пишти **сѣште**. въ църхѣ сѣтъ. (CM, Luke 7:25, 40334)
 idou **hoi** en himatismō endoxō kai trufē **huparkhontes** en tois basileiois eisin (20561)
 Behold, **those who are dressed** in splendid clothing and live in luxury are in kings' courts.

The ACC-GEN annotation is dubious here, since it supposes that “those who are dressed in splendid clothing” refers to some known individuals, i.e. specific rich and powerful people that everyone are familiar with to some degree. Since it is really a general statement about rich people being in kings' courts, KIND, for which we expect LF, is a candidate. However, this is also a special case since it features *iže* acting as a “mock article”⁹⁰. Usually, this variant of *iže* is found together with infinitives, mirroring the use of infinitives with articles in Greek. Therefore, this sentence could be explicitly definite, with definiteness expressed by *iže* instead of LF.

5.2.3.2. Strength in the ACC-INF category for nominalizations

The ACC-INF group contains one single example of SF, out of a total of 43. There is a certain amount of LF here that correspond to anarthrous Greek constructions, but these are all (6) exclusively instances of *desnojq* (right) and *šjujq* (left) and of very little interest. To represent the LF in this group there is the following example:

- (121) Паѣы подобѣно естъ църствие нѣбское неводу въврѣженоу въ море. ꙗ отъ
 въсѣкого рода избѣравѣшю. иже егда исплѣни сѣ извлѣкъше и на краи. ꙗ сѣдѣше
 избраша **добрыѣ** въ сѣсѣды. а **злыѣ** изврѣж вонѣ (CM, Matt 13:47-48, 38841
 & 38842)
 Again, the kingdom of heaven is like a net that was thrown into the sea and gathered
 fish of every kind. When it was full, men drew it ashore and sat down and sorted
the good into containers but threw away **the bad**.

In (121), “the good” and “the bad” is linked back to “kind” in the previous sentence. The one unexpected SF is the following:

- (122) ...а **прочи** имѣше рабѣы его досадиша имѣ и избиша ѿ. (CM, Matt 22:6, 39343)
 ...**hoi de loipoi** kratēsantes tous doulous autou hubrisan kai apekteinan (15762)
 ...while **the rest** seized his servants, treated them shamefully, and killed them.

⁹⁰ In these instances, it is tagged as a demonstrative pronoun. There are only 12 such instances in the corpus: <http://foni.uio.no:3000/lemmata/65290>.

“The rest”, inferred from a previous mention of “they”, should possibly display LF. Since the word directly following *proči* begins with an “i”, this could be an instance of “graphemic interference”. Another possibility is that the OCS translator interpreted it as “others”, i.e. an indefinite amount of people (but not all) of those that are left. With the only SF among 43 being no great mystery, and OCS following article use in Greek except for the expressions mentioned, LF is doubtless the norm.

5.2.3.3. Strength in the ACC-SIT category for nominalizations

The smaller ACC-SIT group, with only 13 sentences, shows more varied use of SF and LF: 4 vs. 9. We would expect LF here as well, as in (123), where the speaker is Jesus. He is addressing the man whose eyes he opened regarding who the Son of Man is, i.e. who he himself is. “He who is speaking to you” is clear from the context, and therefore it has been tagged ACC-SIT:

- (123) видѣлы и еси. ꙗ҃ко ѿ **ѿлаи** съ тобоѡ тѣ естъ. (CM, John 9:37, 42310)
 kai heōrakas auton, kai **ho lalōn** meta sou ekeinos estin (22637)
 You have seen him, and it is **he who is speaking** to you.
- (124) ...ꙗ҃ исплѣниша сѧ страхомъ ѿлѣжаше. ꙗ҃ко видѣхомъ **дивъна** дънесь (CM, Luke 5:26, 51296)
 ...kai eplēsthēsan fobou legontes hoti eidomen **paradoxa** sēmeron (48356)
 ...and were filled with awe, saying, “We have seen **extraordinary things** today.”

The SF among the ACC-SIT show three instances of *grędy* and one *divъna*. The three instances of *grędy*, “he who comes”, are identical to example (115) in chapter 5.2.2.2 (page 81), and should all display LF, but not necessarily *divъna*, displayed in (124). In (124), focus is on the fact that they saw things that were extraordinary and not just mundane, not that they saw a certain set of extraordinary things. The Greek original is anarthrous, and all Western translations also use indefinite forms here. The ACC-SIT tag has been used since what is indirectly referred to here is Jesus’s healing of a paralytic, but the reference is not explicitly marked with a demonstrative or article, and the interpretation is therefore non-specific, and SF is justified. In total, the ACC-SIT category contains one example of what could be considered a rightfully short nominalization, and three erroneous SF. Therefore, just as for the attributes, the entire ACC-group shows that accessible referents get LF.

Table 26 LF/SF distribution in the ACC category for nominalizations

| | SF | LF |
|--------------------------|----------|-------------|
| From corpus data | 7 (7.4%) | 87 (92.6%) |
| Dubious cases to exclude | 3 | 0 |
| Final analysis | 4 (4.4%) | 87 (95.6%)* |

Table displaying the original amount of SF/LF in the ACC category and the final percentage arrived at after all the sentences have been examined.

*p-value < 2.2e-16, 95% confidence interval: 89.1%-98.8%

5.2.4. Short and long form in the NON-SPEC category for nominalizations

As a reminder of what is tagged NON-SPEC, it can be briefly said that these are referents that exist inside certain embeddings (such as negation, modality), outside of which they cannot be referred to. See chapter 4.2 for more specific information. The NON-SPEC nominalizations number 93, with 73 adjectives and 20 participles. For this category, participles and adjectives are treated separately since adjectives show identical amounts of SF and LF, whereas participles mostly show LF. The initial LF percentage is much higher than for attributes at 57.0%, but this number will be dramatically reduced to 10.2% after the analysis. As with the NEW category, this percentage looks high, but these 10.2% translate to just 5 nominalizations.

5.2.4.1. NON-SPEC nominalized participles

The participles show a majority of LF with 13 LF as opposed to 6 SF, half of which correspond to Greek anarthrous nominalizations. Singulars show similar numbers of SF (6) and LF (5), whereas plurals show all LF. Among all the LF, 3 are rightly long as they correspond to restrictive relative clauses that define a hypothetical referent, which in each hypothetical context is definite, as in (125). These are accompanied by dependents that are OLD, such as *ѣ* in (125) and *вамъ* in (126), and are anchored.

- (125) нѣсть рабъ болей ꙗ своего. ни ꙗплъ болей **посълавъшааго ꙗ** (CM, John 13:16, 52050)
ouk estin doulos meizōn tou kuriou autou oude apostolos meizōn **tou pempantos** auton (48719)
A servant is not greater than his master, nor is a messenger greater than **the one who sent** him.

One sentence among the three with OLD dependents, (126), stands out in that it probably belongs in the ACC-INF group; *povelěnaago*, “that which was ordered” refers back to previously given orders that are clear from the context.

- (126) онъ же рече къ нимъ ничтоже боле **повелѣнааго вамъ** творите (CM, Luke 3:13, 40081)
mēden pleon para **to diatetagmenon** humin prassete (20302)
“Don’t collect any more than you are required to,” he told them.
OCS literally: He said to them: nothing more (than) **has-been-ordered** (LF) to you do
- (127) ꙗ ишедъше раби ти на пѣти. събѣраша вса ꙗже обрѣтѣ. зѣлыѣ же и добры. ꙗ исплѣниша бракъ **възлежащихъ**. (CM, Matt 22:10, 39348)
kai exelthontes hoi douloi ekeinoi eis tas hodous sunēgagon pantas hosous heuron, ponērouste kai agathous, kai eplēsthē ho numfōn **anakeimenōn** (15767)
And those servants went out into the roads and gathered all whom they found, both bad and good. So the wedding hall was filled with **guests**.

In (127), the guests in question are not definite; they make up an indefinite mass. LF is used on the previous objects though; *зѣлыѣ ꙗ добры* (these are anarthrous in Greek), and this could provoke LF on

vъzležęstichъ, unless, of course, what the translator wanted to say was “filling the wedding with *these* guests”, which justifies LF use. All these LF nominalizations are therefore excluded.

- (128) **Слоушаѣи** васъ мене слоушааѣ. ꙗ **отъмѣтаѣи** сѧ васъ мене сѧ отъмѣтааѣ (CM, Luke 10:16, 40591)
Ho akouōn humōn emou akouei, kai **ho athetōn** humas eme athetei (20832)
The one who hears you hears me, and **the one who rejects** you rejects me.

Sentences such as (128) are quantified without having any explicit quantifier, which would be “all those who” or “every person who”. They have the meaning of “whoever does this or that”, i.e. the nuance a nominalized participle with an article in Greek usually carries (which should be tagged QUANT). Among the 13 LF, there are 5 that are quantified and 5 that are generic (“the poor”, “the rich”, “the hungry”). These are excluded. Real NON-SPEC should be short, whereas quantified referents should be long (cf. table 5 in chapter 4.2, where QUANT belongs in the non-specific camp together with NON-SPEC). I have chosen to exclude referents that can be interpreted as generic or quantified from the final statistics in table 27. In Appendix 1, only the sentences that should *definitely* be tagged KIND and QUANT are found; the borderline cases are not moved to another category, they are merely excluded.

In the small group of 6 (expected) SF, which are all singulars, there are three instances where the referent is “nothing”. The “nothing... which...” construction appears three times in this group. In example (129), of which there are two, we see a typical example of a NON-SPEC singular referent. The last sentence is very similar, also displaying a “as one who” structure with *ěko* but with the verb *služiti* instead of *iměti*.

- (129) бѣ бо оучѧ ѣко власть **ѣмѧ**. ꙗ не ѣко кѣнижѣници ихъ (CM, Mark 1:22, 50222)
ēn gar didaskōn autous hōs exousian **ekhōn**, kai oukh hōs hoi grammateis (6528)
For he taught them as **one who had** authority, and not as the scribes.

For participles, which are not very numerous in the NON-SPEC group, LF predominates in two types of sentences: those that can also have a generic sense and those that are implicitly quantified to include “all” or “every” of possible referents. Such sentences are mainly plural. Sentences that are non-quantified NON-SPEC show SF. As such, SF for NON-SPEC participles is at 100%.

5.2.4.2. NON-SPEC nominalized adjectives

Adjectives show identical SF/LF distribution: 38 LF vs. 38 SF. As will be shown, it is clear that for true NON-SPEC SF is expected. Most LF are actually instances of generic referents.

In the SF group, there are 18 occurrences of adjectives cooccurring with *čьto*, “something”, or *ničьtože*, “nothing”, with many repeating lemmas: *зълъ*, “evil”, *bolii*, “more”, *dobrъ*, “good”, *lichъ*, “bad”, *tainъ*, “secret”. In the PROIEL annotation, in cases where adjectives cooccur with these

indefinite pronouns the adjective is considered to be the head, but there is no strong justification for this choice. If the pronoun would instead be considered the head, these would be attributive adjectives and not nominalizations. A common trait among these adjectives is that they are mainly abstract⁹¹ (good, evil, worthy, etc.). 75% of the adjectives in the group are abstract, whereas 12% denote physical characteristics (small, strong, mute) and another 12% various other types of adjectives. Among the Greek originals there are only two that are arthrous. Furthermore, 84.9% are in the singular and 87.8% are neutral. As usual with non-specific referents, there are several instances where a non-specific and generic interpretation makes very little difference, but also some sentences that clearly do demonstrate the difference, as does the following:

- (130) алчѡштаѡ исплѣни **благѣ**. ꙗ богаташтаѡ сѧ отъпусти тѣшта. (СМ, Luke 1:53, 51237)
 peinōntas eneplēsen **agathōn** kai ploutountas exapesteilen kenous (48342)
 He has filled the hungry with **good things**, and the rich he has sent away empty.

In (130), the non-specific “good things” is gen.pl. SF, whereas the generic “the hungry” and “the rich” show LF. The rest of the SF all make sense and there are no erroneous SF uses among these 33.

In the LF group, we find 80.0% plurals and 77.5% masculine, no occurrences of *čьto* or *ničьtože*, and a roughly equal distribution of lemmas denoting physical characteristics and abstract qualities. A fourth are arthrous in Greek, however it is clear in this group, where many of the sentences come from the same biblical passage, that the translator saw what was non-specific in Greek as generic, and therefore added the LF. There are numerous instances of “blind”, “poor”, etc., that are all easily interpreted as generic, as in (131).

- (131) слѣпѣи прозираѡтъ. ꙗ хромии ходѡтъ. прокажении очищаѡтъ сѧ и глоусии слышѡтъ. мрѣтвиѣ вѣстаѡтъ. ꙗ ништиѣ благовѣстоуѡтъ. (СМ, Matt 11:5, 38658)
 tufloi anablepousin kai **khōloi** peripatousin, **leproi** katharizontai kai **kōfoi** akouousin, kai **nekroi** egeirontai kai **ptōkhai** euangelizontai (15073)
 The blind receive their sight and the lame walk, lepers are cleansed and the deaf hear, and the dead are raised up, and the poor have good news preached to them.

Once all the generic instances have been removed, there are only 6 LF left that are quite likely non-specific, among which some are probably not supposed to be long. 2 correspond to arthrous Greek sentences, and are found in sentence (132) below, featuring good/bad. LF could therefore be “blamed” on the Greek, since there is no other explanation for it. The original meaning seems to be something like “that which is good/evil”, but the annotators working on the Greek text apparently saw this as non-specific rather than a generic referent.

⁹¹ To see the split between abstract/non-abstract, see Appendix 2.

- (132) ...и изиджъ сътворишеи **благаа** въ въскръшение животоу. а сътворишеи **зълаа** въ въскръшение сждоу (CM, John 5:29, 41890)
 kai ekporeusontai hoi **ta agatha** poiēsantes eis anastasin zōēs, hoi **ta faula** praxantes eis anastasin kriseōs (59959)
 ...and come out, those who have done **good** to the resurrection of life, and those who have done **evil** to the resurrection of judgment.
- (133) аще оубо въ лжкавъни сжште. оумѣете дааниѣ блага даѣти чадомъ вашимъ. кольми паче ѿтцъ вашъ иже естъ на небесехъ. дасть **благаа** просащимъ его (CM, Matt 7:11, 38471)
 ei oun humeis ponēroi ontes oidate domata agatha didonai tois teknois humōn, posō mallon ho patēr humōn ho en tois ouranois dōsei **agatha** tois aitousin auton (14885)
 If you then, who are evil, know how to give good gifts to your children, how much more will your Father who is in heaven give **good things** to those who ask him!

There is another instance of LF *blagaa* in (133) where a SF would be expected, but without an arthrous Greek original. It is possible that the OCS translator saw this *blagaa* as referring back to the first SF *daanijě blaga* (underlined in the example). The same phenomenon could be the explanation for the LF in (134). Here, “new wine” could have been interpreted as OLD by the translator since there has been repeated mention of “new wine” in the text preceding this line, but no previous mention of old wine, only “old wineskins”, which would explain why one displays SF and the other LF. As such, this sentence is excluded based on its probably being interpreted as OLD by the OCS translator. None of the other 5 LF will be excluded.

- (134) ꙗкоже пивъ ветѣха абие хоштеть **новоуоумоу** (CM, Luke 5:39, 40219)
 kai oudeis piōn palaion thelei **neon** (20443)
 And no one after drinking old wine desires **new**.
- (135) по **тоуждемъ** же не иджъ. нъ бѣжатъ отъ него. ѣко не знажъ тоуждиихъ гласа. (CM, John 10:5, 42324)
 allotriō de ou mē akolouthēsousin, alla feuxontai ap’ autou, hoti ouk oidasin tōn allotriōn tēn fōnēn (22653)
A stranger they will not follow, but they will flee from him, for they do not know the voice of strangers.

The LF in (135) is somewhat more difficult to explain. The stranger in question is hypothetical and non-referential, and unless this is a narrative device similar to the one displayed in (105) (page 76), LF must be considered an error. This example will not be excluded. Of the 5 LF that remain as “unexpected”, 4/5 are neutral plural of the *blagaa* type.

All in all, the percentage of LF among the nominalized adjectives is at 13.2%. Some of the remaining LF could possibly be excluded and lower this percentage. With adjectives and participles combined, the percentage is smaller, as is seen in table 27. The total number of actual non-specific referents is

rather small as well, since very many sentences are actually generic or quantified in nature. There is no statistically significant difference for NON-SPEC attributes and nominalizations⁹².

Table 27 LF/SF distribution in the NON-SPEC category for nominalizations

| | SF | LF |
|--------------------------|-------------|------------|
| From corpus data | 44 (45.8%) | 52 (54.2%) |
| Dubious cases to exclude | 0 | 46 |
| Final analysis | 44 (89.8%)* | 5 (10.2%) |

Table displaying the original amount of SF/LF in the NON-SPEC category and the final percentage arrived at after all the sentences have been examined. Many exclusions are generic or quantified, see Appendix 1.
p-value = 7.597e-09, 95% confidence interval: 77.8%-96.6% (using a binomial test)

5.2.5. Short and long form in the KIND category for nominalizations

Generic nominalizations make up a total of 80 participles or adjectives that refer to kinds. LF clearly dominates with 83.8% in this group, which contains mostly plurals (77.5%). However, this group should really be larger if all generic LF nominalizations found in other categories were added. The fact that these generic referents (including referents erroneously categorized as NON-SPEC) together with the quantified examined in the next sub-chapter, are rather numerous, and also predominantly plural, could be part of the explanation for Flier's view that LF is more frequent among plurals.

Table 28 Number and strength for nominalized generics

| | Short form | Long form | Total |
|-------|------------|------------|------------|
| sg | 6 | 12 | 18 (22.5%) |
| pl | 7 | 55 | 62 (77.5%) |
| Total | 13 (16.3%) | 67 (83.7%) | 80 (100%) |

There is no doubt that nominalized generics strongly tend to display LF, which is made all the more clear by the fact that the ones displaying SF make out a small amount of sentences that are repeated throughout the codices. One of the sentences (136) contains all of 4 coordinated⁹³ SF that should have displayed LF if speaking of "the good/evil/etc." in general:

- (136) ...да бждете ѿнве ѿца вашего иже естъ на Ѡбсхъ. Ѡко слъньце свое
съѣать на **зълы** и **благы**. и дъждитъ на **праведъны** и на **неправедъны**. (CM,
Matt 5:45, 59990)
...hopōs genēsthe huiōi tou patros humōn tou en ouranois, hoti ton hēlion autou

⁹² P-value: 1.0 using a 2x2 contingency table and a Fisher's exact test.

⁹³ Because of the habit in OCS to only mark the first conjunct in a series of coordinated attributes with LF, second conjuncts are excluded, and therefore *blagy* in (136) is not part of the results given by the script. See chapter 4.3.3.5.

- anatellei epi **ponērous** kai **agathous** kai brekhei epi **dikaious** kai **adikous** (47700)
 ...that you may be children of your Father in heaven. He causes his sun to rise on **the evil** and **the good**, and sends rain on **the righteous** and **the unrighteous**.
 (137) **исповѣдаѣ ти сѧ ѿче ѿи нѣсе и землѧ. ѣко оутаилѣ еси се ѿтъ прѣмждрѣ и разоумѣнѣ. и ѿткры си млaddenъцемѣ.**(CM, Luke 10:21, 40599)
 exomologoumai soi pater, kurie tou ouranou kai tēs gēs, hoti apekrufas tauta apo **sofōn** kai **sunetōn**, kai apekalufas auta nēpiois (20841)
 I praise you, Father, Lord of heaven and earth, because you have hidden these things from **the wise** and **learned**, and revealed them to little children.

(136) comes across as a generic sentence (about the sun) with generic referents (the good etc.), which show article use in English as well as French and Danish, and lack of article in Greek, Swedish and Norwegian. A non-specific reading, referring to non-specific good and evil people, is also possible, even though the *statement* is generic. If we posit that OCS should show LF for generics, then this has probably been interpreted as non-specific based on the anarthrous Greek. Therefore, it is difficult to outright say if this is an erroneous SF use or not. (137) also seems anomalous based on the English translation, unless we accord it a non-specific reading (“from wise and learned people”), and just like the previous example it also corresponds to an anarthrous Greek sentence. For the attributive group, 69.7% (23)⁹⁴ of the LF corresponded to arthrous Greek constructions, and for the nominalizations this percentage is very similar at 64.5% (40). These examples illustrate the difficulty in differing between KIND and NON-SPEC. Often it becomes a question of letting the use of SF/LF dictate the OCS interpretation.

Another exception that appears three times throughout the corpus is the classic displayed below:

- (138) **оудобѣе естъ вельбѣдоу сквозѣ игълинѣ оуши проити. неже богатоу въ цѣрствіе бжіе вѣнати** (CM, Mark 10:25, 36923)
 eukopōteron estin kamēlon dia tēs trumalias tēs rafidos dielthein ē **plousion** eis tēn basileian tou theou eiselthein (7086)
 It is easier for a camel to go through the eye of a needle than for **a rich person** to enter the kingdom of God.

This sentence also illustrates the problem of deciding whether a singular is non-specific or generic. Greek, English, French, and all of the Scandinavian languages use indefinite constructions for the rich person referred to in this sentence, and it would seem plausible to interpret this as non-specific. The remaining sentences displaying SF are more examples of difficult cases of singulars that could be interpreted either as KIND or NON-SPEC.

Four more SF come from the same sentence, speaking of the “a little” and “a lot” mentioned in example (114) (page 80) among the OLD nominalizations. It is dubious whether this belongs in the

⁹⁴ 2 out of 35 KIND attributes and 5 of the 80 KIND nominalizations stem from CS, which has no Greek originals in the corpus. These are not included when percentages for arthrous Greek originals are calculated.

KIND category. If there is no referent, the tag should probably be NON-SPEC, and these are therefore excluded from the final analysis. Another exclusion from the KIND group is something that could also be interpreted as non-specific (i.e. “in a dry...”):

- (139) зане аште въ сырѣ дрѣвѣ си творятъ. въ **соусѣ** что бждеть (CM, Luke 23:31, 41494)
hoti ei en tō hugrō xulō tauta poiouein, en **tō xērō** ti genētai; (21771)
For if they do these things when the wood is green, what will happen when it is dry?”
Literally: ...in **dry** what will be

It is possible that even more of the SF actually belong in the non-specific category, so the percentage for SF in table 29 may still be too high.

Table 29 LF/SF distribution in the KIND category for nominalizations

| | SF | LF |
|--------------------------|------------|-------------|
| From corpus data | 13 (16.3%) | 67 (83.7%) |
| Dubious cases to exclude | 8 | 0 |
| Final analysis | 5 (6.9%) | 67 (93.1%)* |

Table displaying the original amount of SF/LF in the KIND category and the final percentage arrived at after all the sentences have been examined. Exclusions include things that belong in the NON-SPEC or QUANT categories. Sentences can be found in Appendix 1.

*p-value = 6.388e-15, 95% confidence interval: 84.5%-97.7%, (using a binomial test)

5.2.6. Short and long form in the QUANT category for nominalizations

QUANT is the tag used for non-specific referents that are found with quantifier restrictions, decided by formal criteria (see below). The nominalized QUANTs are very numerous with a total of 249. This group contains a great number of overtly quantifying words (*всѣхъ*, *многъ*, *всѣ*) as heads. Since these quantifiers behave more like pronouns than adjectives and show very little variation in strength⁹⁵, they will not be included in the final statistics for SF/LF, meaning that most of the quantified nominalizations that remain are participles. They are *not* excluded when cooccurring with a nominalization though. That mostly participles remain is not surprising since the annotation of QUANT in Greek, which is where the majority of the tags come from, is based on formal criteria: headless relative clauses, overt quantifiers (i.e. Greek words corresponding to *всѣхъ*, *многъ*, *всѣ*) and nominalized participles with a definite article are given this tag. There is a clear division of SF and LF between adjectives (*всѣхъ* and *многъ*, which will be excluded below) and participles, indicating that what we should really expect here is LF, which is also confirmed by the LF percentage of 92.2% in table 30.

⁹⁵ These lemmas, with the exclusion of *всѣ*, are discussed in chapter 4.3.3.2.

Table 30 Distribution of SF/LF among participles and adjectives in the QUANT category

| | Part | Adj |
|-------|--------------|-------------|
| SF | 8 (5.1%) | 78 (84.8%) |
| LF | 149 (94.9%) | 14 (15.2%) |
| Total | 157 (100.0%) | 92 (100.0%) |

The distribution appears rather clear-cut, but the apparent exceptions will be examined. The quantifiers *мѣногѣ* and *вѣсѣкѣ* (which will not count in the final analysis) stand for most of the SF with only 2 LF among them. (140) illustrates a typical example of a quantified nominalization with a nominalized participle.

- (140) **ѿлаи** о себѣ славы своеѣ иштеть (CM, John 7:18, 42049)
ho af' heautou **lalōn** tēn doxan tēn idian zētei (22350)
Whoever speaks on their own does so to gain personal glory.

Once *вѣсѣкѣ*, *мѣногѣ* and *вѣсѣ* have been removed as *heads*, 170 nominalizations remain: 151 participles and 18 adjectives, with a slight majority of singulars. In the small group of 18 adjectives, the SF/LF distribution is rather equal. There are 11 that cooccur with one of the three quantifiers. 6 are *мѣногѣ* + adjective, which can very well rightly show SF, and the LF are not unexpected either for any of the three cooccurring quantifiers. This leaves 7 “independent” adjectival nominalizations, of which only one is short. When these correspond to Greek adjectives, a mistake has been made by the annotator, since nominalized adjectives are not to be tagged QUANT. Otherwise, they are mismatches between Greek and OCS structures. Either way, the tagging should be changed.

The participles (151) show a very small number of SF. The sheer number of LF makes any extensive comment on this group superfluous, and LF can safely be established as the norm. There are a total of 7 SF, of which one is *sy*, which shows vacillation, and one is a resultative past participle in -l, which have no LF. These are excluded. Some are understandably short, notably when cooccurring with *мѣногѣ* as in “many who had various diseases”, which makes possible an explicit difference in meaning compared to use with LF, “the many bearing various diseases”.

The QUANT group as a whole displays both SF and LF. In cooccurrence with *вѣсѣ* and *вѣсѣкѣ*⁹⁶, LF is expected, but for *мѣногѣ*, both LF and SF can be explained, as “many of X” (SF) or “the many X” (LF). It would seem like OCS makes use of both variants, since both SF and LF are found in cooccurrence with *мѣногѣ*. In the case of bare nominalized participles, the mere LF marks the expression as

⁹⁶ This is true for *participles* cooccurring with *вѣсѣкѣ*. Flier (1974, p. 141) only finds five occurrences of *вѣсѣкѣ* cooccurring with an adjective in the Gospel texts, of which all except one are short.

quantifying unless the context says otherwise. With such an analysis, LF could be said to be the “normal” expression of quantification, with a large initial amount of SF due to the explicit markers acting like grammaticalized units.

Table 31 LF/SF distribution in the QUANT category for nominalizations

| | SF | LF |
|---|------------|--------------|
| From corpus data | 86 (34.5%) | 163 (65.5%) |
| Excluded SF-only lemmas & dubious cases | 73 | 9 |
| Final analysis | 13 (7.8%) | 154 (92.2%)* |

Table displaying the original amount of SF/LF in the QUANT category and the final percentage arrived at after all the sentences have been examined. Exclusions are mainly lemmas that only display SF, i.e. overt quantifiers.

*p-value < 2.2e-16, 95% confidence interval: 87.1%-95.8%

5.2.7. Conclusion regarding nominalizations

Nominalizations have been claimed to show an unexpected amount of LF, especially in the plural, and this has then been seen as the beginning of the fall of the SF/LF distinction as a system of definiteness marking, notably by Michael Flier. The data from the PROIEL corpus, combining SF/LF with IS show that instances of LF are usually motivated by semantics or pragmatics. If one merely looks at the number of LF among the nominalizations, which gives a percentage of 76.3%, the conclusions of previous research seem correct. However, as can be seen from the “Final analysis” percentages in the tables presented in chapter 5.2, the LF percentage drastically decreases in unexpected categories once analyzed. The “error margin” is therefore not very much greater for nominalizations than it is for attributes. As is seen from the numbers presented in chapter 6, there are no significant differences between attributes and nominalizations. The NEW category at first glance seems to contradict this statement, but taking the numbers and sentences concerned into consideration, this category actually does not show any large amount of LF. Furthermore, we have no indications that there is any “overuse” of LF in younger texts, such as CS. The scarcity of the material available from CS at the time of the data extraction could be an issue; there are only 55 nominalizations in total from CS, and these actually show *less* LF, at 67.3%. However, Rondestvedt, who has examined somewhat younger original Slavic texts (i.e. not translations from Greek), found no results indicating such a trend. Flier’s impression regarding LF nominalizations being overly frequent is probably explained by his decision to only view singulars as generic.

5.3. Appositions

Following Meyer (1992, p. 24), an apposition is a grammatical relation that can have various kinds of syntactic realizations. The simplest example of this would be the type represented by (141).

(141) He met *Anne, a French girl*, at the party.

This sentence represents the archetypical apposition, that is a juxtaposition of two co-referential noun phrases. Appositions can be many other things though, and what is actually counted as an apposition can vary from definition to definition. Since we are dealing with adjectives and participles, our appositions will not be nouns, but mostly nominalizations that ascribe properties or actions to a head word. The adjectives can themselves be derived from nouns.

What kind of behavior is expected in the SF/LF dichotomy for appositions? One would perhaps expect to find a majority of LF since the reference could be considered as already mentioned and therefore as known. On the other hand, Meyer (1992, p. 20) lists the pragmatic characteristics of appositions as “new or partially new information”, bringing to mind predicative expressions. That Anne in (141) is French is after all a new piece of information. One could, of course, imagine that this is already known, and that Anne would be referred to as “Anne, the French girl”. This is an important split between qualifying versus identifying appositions, the same split that is found for predicatives. For identifying appositions, *both* the head and the apposition are referential, both referring to something/someone that should be already known. A qualifying apposition can be referring as well, but to something that is new or non-specific. Unfortunately, appositions get no IS-status of their own, so they cannot be separated from the qualifying appositions automatically. We would, however, expect LF on the identifying and SF for qualifying appositions.

Meyer (1992, p. 32) further states that in English, appositions where the first unit is indefinite and the second unit is definite, or vice versa, are uncommon. We will find such instances in section 5.3.2 however. Since nominalizations have so far not been found to use more LF than the pragmatics would have them do we shouldn't expect LF purely on the basis of the adjectives/participles being nominalizations. Is there any difference in behavior when the head word itself is an apposition on something else? Do adjectives and participles behave differently? In this chapter, nominalized appositions on proper nouns will first be examined, then attributes to appositions and finally other nominalized appositions.

5.3.1. Nominalized appositions on proper nouns

Undoubtedly, from a pragmatic point of view, adjectives and participles that are appositions to proper nouns should be long, unless other factors are involved, such as suffixes that show only SF. Flier (1974, p. 85) clearly says that LF is always used in instances when the person in question is

associated with a place using an -ъsk adjective, but since an individual that is referred to by a proper name is usually known, so should a clear majority of other appositions as long as they are identifying. This is a group that can easily be checked, and as will become clear, LF is really the norm.

Among the appositions present in the IS corpus, 64 have a proper noun as their head word. A further division into adjectives and participles gives us 33 of the former and 31 of the latter, and an overall LF percentage of 95.3%. LF is represented by example (142) below, featuring an epithet that equals Paul to “the very simple (one)”.

- (142) живъ же єдинъ о себѣ пауль **прѣпростыи**. лѣто єдино. (CS, Paul the Simple, 58692)
Paul **the very simple** lived there alone by himself for one year.⁹⁷

The adjectives show no variation and are all long. What is perhaps the most notable feature in this group is that 42.4% stem from CS, which only makes up 8.1% of the IS corpus. The IS of these proper nouns show a majority of OLD and ACCESSIBLE tags, with two erroneous NEW-tags denoting well-known referents. It is also interesting to note that among these 33 adjectives, 13 are formed with the -ъsk suffix: *iskariotъskъ* (9), *pontъskъ* (2), *galileiskъ* (1) and *syrъskъ* (1). Most of the referents are the well-known Judas Iscariot and Pontius Pilate. We saw in chapter 5.1.1.1 that -ъsk adjectives formed from places and peoples behave erratically, but they do not show unexpected forms in this group. The tendency to use LF on adjectives used as appositions on proper nouns therefore seems very strong. Word order seems to be of no importance: 15 of the adjectives precede their heads, whereas 21 follow it, which shows an unusual amount of pre-nominal modifiers at 41.7%.

When it comes to participles, we are back to a more proportionate amount of results from CS (3.2%), but the amount of SF is higher: 16.1% (5). There is also a radical difference in word order: none of the participles precedes its head.

Among the SF participles, some can be discarded straight away: one is the participle *sy*, which shows some odd vacillation (see chapter 4.3.3.4); another is *sqštejq* (instr.sg.f of *sy*), which is ambiguous when it comes to SF/LF, (see table 2 in chapter 3.1). Another SF cooccurs with the specific *jedinъ*, “a certain”, and the head has been tagged as NEW. The presence of *jedinъ* is very likely what provokes SF in the participle. There are 2 more SF. These are found in somewhat peculiar sentences:

- (143) приде бо иоанъ не **пих** ни ѣды. и їлѣтъ (CM, Matt 11:18, 50801)
 ēlthen gar īōannēs mēte esthiōn mēte **pinōn**, kai legousin (47832)
 For John came neither eating nor **drinking**, and they say...
 (144) ѣкоже естъ писано въ кѣнигахъ словесъ. исаиѣ прка їлѣшта (CM, Luke 3:4, 40061)
 hōs gegraptai en biblō logon Hēsaïou tou profōtou (20282)

⁹⁷ Translation from: http://clover.slavic.pitt.edu:8080/exist/paul/data/paul_main.html (2013-05-07)

As it is written in the book of the words of Isaiah the prophet ...

OCS literally: As is written in the books of the words of Isaiah, the prophet, **saying**...

In (144) the Greek sentence contains no additional “saying”, *glōssa*, which is an addition from the OCS translator. As an apposition to Isaiah, it should be long, especially since the Greek has an article on “prophet”, making it quite improbable that the OCS should read “Isaiah, a prophet, who says”. However, as has been seen earlier with identical sentences, the participle could simply be adverbial and therefore short. In “neither eating nor drinking”, example (143), the analysis where the participles are appositions to John competes with a purely adverbial analysis. Seen as adverbs, they are modifiers to the verb “arrived”. Since the fact that John is “neither eating nor drinking” is a constant state, not something directly related to the fact that he arrived, the appositive analysis won out. This ambiguity could, on the other hand, explain the use of SF.

In conclusion, the use of LF for appositions on proper nouns appears to be a very strong trend, corresponding to the idea that LF expresses definiteness in the form of identifiability. The total percentage of SF for appositions on proper nouns is 4.6%⁹⁸ (3).

5.3.2. Attributes to appositions

When the head word (nouns only) is itself an apposition, the form of the modifying expression should logically follow the definiteness of the head word of the entire expression. If Jesus is the head word, then should not the apposition and its attribute be definite, since he is known by all? This is not necessarily the case, and it is easy to see why not. Looking at the data from the corpus, we find 35 instances of appositive head words, i.e. an appositive noun modified by an adjective (26) or a participle (9). The strength of the apposition will be shown to correspond to the information status of the head, with the exception of non-referring appositions functioning as something akin to qualifying predicative expressions (examples below). We will first deal with the possibly “unexpected” SF (12) and then the LF (23). In the examples, the head word is underlined and the apposition and its attribute is bold.

In the group of SF attributes on appositions (12) we find several describing, qualifying expressions, where the apposition bestows a quality upon the head word, and all but one are adjectives. The SF is then fully expected, since when ascribing a quality of the type “Jesus, a good man”, the person or thing in question (Jesus) is given the quality of any good man, not a specific, known good man. It is similar to a predicative expression of the kind “Jesus is a good man”. (145), (146) and (147) show examples of two clearly known personas and one new being introduced with indefinite appositions:

⁹⁸ P-value = 4.743e-15 (using a binomial test).

- (145) Приде иосифъ отъ ариматѣѡ. **благообразенъ съвѣтъникъ**. (CM, Mark 15:43, 37335)
 elthōn lōsēf ho apo Arimathaias **euskhēmōn bouleutēs** (7513)
Joseph of Arimathaea, an **honourable counselor**.
- (146) Ёже о ѣсѣ назарѣнинъ. иже бысть мѡжъ **пѣркъ силенъ** (CM, Luke 24:19, 41550)
 ta peri lēsou tou Nazarēnou, hos egeneto anēr **profētēs dunatos** en ergō (21829)
 Concerning Jesus of Nazareth, which was **a prophet mighty in deed**.
- (147) Бѣ же ѡлвкъ отъ фарисѣи. никодимъ имѡ емоу. **кѡнасъ иудеискъ**. (CM, John 3:1, 41703)
 ēn de anthrōpos ek tōn Farisaiōn, Nikodēmos onoma autō, **archon tōn loudaiōn** (21998)
 There was a man of the Pharisees named Nicodemus, **a ruler of the Jews**.

All of these examples show indefinite article in English and lack articles in Greek. The use of SF here therefore clearly seems to correspond to non-specific indefiniteness or non-referentiality. Other SF examples feature the concept *oko lqkavъno*, “envy”, and the indefinite “a treasure in the heavens”. For neither would we expect LF. Other sentences found in this group either display behavior of the kind exemplified in (145)–(147) or include the -ъsk suffix for peoples and places (discussed in chapter 5.1.1.3), notably *ijudeiskъ* (6 occurrences), which was shown to almost exclusively display SF, and need not be examined in detail here.

Among the 23 LF we find many well-known referents. The referents in this group are Apollon (1), Christ (5), God (4), the Holy Spirit (3) and Paul (2), who are all clearly definite and who are further marked as such with definite modifiers. The opposite also occurs: the head is either specific or non-specific indefinite, and the apposition specifies the referent. In (148) we see “the Spirit of truth” (the Holy Spirit) analyzed as an apposition to the specific (or non-specific) “another helper” (analyzed as NEW in the Greek text):

- (148) и азъ оумолѣѡ бѣтца. и иного параклита дасть вамъ. да бѣдетъ съ вами въ вѣкъ.
дѣхъ истинъны. егоже миръ не можетъ приѣти. ѣко не видитъ его ни знаѡтъ его.
 (CM, John 14:16-17, 42640)
 kai egō erōtēsō ton patera kai allon paraklēton dōsei humin, hina meth’ humōn ē eis ton aiōna, **to pneuma tēs alētheias**, ho ho kosmos ou dunatai labein, hoti ou theōrei auto oude ginōskei auto (22980)
 And I will pray the Father, and He will give you another Helper, that He may abide with you forever— **the Spirit of truth**, whom the world cannot receive, because it neither sees Him nor knows Him.

“The Spirit of Truth”⁹⁹ displayed above appears three times in this group of LF appositions. We also find the above-mentioned suffixes -n and -ъsk; *istinъnъ* (4) and *člověčъskъ* (2). As mentioned by Flier

⁹⁹ The OCS apposition here does not follow the rule of case agreement here. The head word is in the genitive-formed accusative, whereas the apposition is in the nominative-formed accusative, possibly due to hesitation between N- and G-formed accusative for animate direct objects.

(1974, p. 104), *istinъnъ*, “true”, modifies unique referents (“the one and only”, “the true one”) and is expected to always be long (but is actually more often *short*¹⁰⁰).

With participles it is not a question of comparison or ascribing a quality, but of equaling the head word with the highly specific “the one who + verbal action”, as in, “Mary, **(the one) who was promised** to him as wife”, i.e. a nominalization. Another example:

- (149) азъ вѣровахъ ѣко ты еси хъ снъ бжіи. **градъи** въ весь миръ. (СМ, John 11:27, 42426)
egō pepisteuka hoti su ei ho Khristos **ho huios** tou theou **ho** eis ton kosmon **erkhomenos** (22756)
I believe that thou art the Christ, **the Son** of God, **which should come** into the world.

Adjectives and participles whose heads are appositions therefore seem to behave as expected, displaying LF for familiar and specific items and SF for those that are non-specific or non-referential. That is, they follow the general rule posited for the SF/LF distinction.

5.3.3. Nominalized appositions on common nouns

This group of 25 appositions contains 24 participles and 1 adjective that are appositions on common nouns¹⁰¹. These 25 appositions are mainly LF nominalizations, but there are also 6 SF. The LF will be looked at first, then the SF. In the examples, the nominalization is in bold, and the noun it is an apposition to is underlined.

In the LF group, the IS-tags of the head nouns are mainly what we expect: OLD, ACC, KIND. These are represented by example (150), as indeed very many of them have “God” as head.

- (150) видѣвъше же народи чюдиша сѧ и прославиша ба. **давъшааго** власть такъ
ѣлвкомъ. (СМ, MATT 9:8, 38551)
idontes de hoi okhloi efobēsēsan kai edoxasan ton theon **ton donta** exousian
toiautēn tois anthrōpois (14964)
When the crowds saw it, they were afraid, and they glorified God, **who had given**
such authority to men.

There are, however, two heads tagged NEW among the LF. These are examples of a type we have already encountered in the NEW chapter, where a referent is introduced (“a woman”, “a prisoner”) and then identified (“...the one named Litus/Barabbas”). Since these appositions are identifying and not qualifying, LF is expected. They should, however, probably have been annotated as attributes, not appositions. Both of these examples are given below.

¹⁰⁰ 17/32 occurrences in the PROIEL corpus are short, i.e. 53.0%.

¹⁰¹ By narrowing it down to common nouns, many appositions are excluded. Appositions on demonstrative pronouns (18) and personal pronouns (11) are frequent, but for these we expect and find uniquely LF, so they are not discussed there

- (151) Сий же иже близъ дия. глаголемый аполонъ. не отъ блжда ли роди сѧ. отъ жены (NEW) **рекомыѧ** литоус. (CS, Vita of Paul and Juliana, 81056)
This one, who is close to Zeus, the one called Appolon, was he not born from debauchery, from a woman **named** Litus?
- (152) имѡаше же тѣгда сѡвѣзѣнѣ (NEW) нарочита **нарицаемааго** варава. (CM, Matt 27:16, 39738)
eikhon de tote desmion episēmon, **legomenon** Barabban (16161)
At that time they had a well-known prisoner **whose name was** Barabbas.

In the SF group (6), half of the IS-tags are of the kind for which LF is expected: ACC-GEN (1), OLD (2), but there are also NEW (2) and NON-SPEC (1). The SF NEW is presented below. The NON-SPEC sentence features a resultative I-participle, which do not make use of the SF/LF distinction.

- (153) ꙗ се мжжѣ (NEW) именемъ иосифѣ. сѡвѣтъникъ **сѣ**. мжжѣ благъ и правѣденъ. (CM, Luke 23:50, 51745)
kai idou anēr onomati iōsēf bouleutēs **huparkhōn**, kai anēr agathos kai dikaios (48599)
Now there was a man named Joseph, from the Jewish town of Arimathea. He **was** a member of the council, a good and righteous man.
OCS literally: And behold a man named Joseph, a councillor, **being** a good and righteous man.
- (154) Въ тѣ день пристѣпиша къ нему ѣлжште садоукеи (NEW) (CM, Matt 22:23, 39372)
en ekeinē tē hēmera prosēlthon auto Zaddoukaioi, **legontes** (15792)
The same day Sadducees came to him, **who say** that...

In (153) this particular participle, *sy*, has previously shown some unexpected behavior, but in this case we would expect a SF since this is not an identifying apposition, but a qualifying one. In (154), the Sadducees are clearly indefinite, and the SF on the participle is expected.

Among the SF nominalized appositions on OLD or ACC-GEN referents, we find a duplicate of the structure presented in example (143) (chapter 5.3.1, page 95), with the only difference that this time, it is the Son of Man who is eating and drinking. SF is justified in the same way for the Son of Man as for John. There is also another example featuring the participle *glagoliqštemъ*, which could be adverbial (cf. examples (74), page 61 and (144), page 95).

5.3.4. Conclusion regarding appositions

As a group, appositions seemingly do not behave differently from attributes. When they are referential, they display LF for known referents and SF for unknown, and when non-referential, they display SF. In the instances where we find mismatches between IS-status and expected strength of the adjective or participle there is always an explanation, meaning that the appositions display very few erroneous forms.

5.4. Vocatives

Vocatives as a group (59) have not been included in the other analysis chapters since there is general agreement among Lunt, Flier, Borodič and more or less everyone that has written on the subject that vocatives are expected to appear in the LF. When addressing someone, the speaker will, in general, have a certain person (or several people) in mind, or, as Flier (1974, p. 152) puts it: “Such referents are obviously definite”. Vocatives usually have the IS-tag OLD. Nonetheless, they *need* not be specific; you can very well address a generic group of people you do not know anything about (“O sinful women!”). Among the nominalized adjectives and participles in the vocative in the corpus, there is one such example:

- (155) горе вамъ **смѣѣшѣи сѧ** (QUANT) нынѣ ѣко въздрыдаате и възплачете сѧ (СМ,
Luke 6:25, 40252)
ouai, **hoi gelōntes** nun, hoti penthēsete kai klausetē (20479)
Woe to you **who laugh** now, for you shall mourn and weep.

Among the world’s languages, vocatives are generally unmarked for definiteness, but for example Serbo-Croat marks the accompanying adjective (Lyons, 1999, p. 153). This is referred to by some as “redundant marking” (see chapter 2.2.1), and NT Greek, which frequently uses articles for proper nouns, does not use them for vocatives (Flier, 1974, p. 152). If OCS is supposed to mark vocatives as definite, it may appear a bit surprising that among the nominalizations the majority are actually short, as can be seen in table 32.

Table 32 Vocatives and the distribution of short/long form

| | SF | LF |
|-----------------|------------|------------|
| Nominalizations | 14 (58.3%) | 10 (41.7%) |
| Attributes | 7 (20.0%) | 28 (80.0%) |

Flier (1974, p. 152) suggests that in cases of SF for OCS vocatives, there is Greek influence. Lunt (2001, p. 142), on the other hand, argues that there are two cases where vocatives will be in the SF:

- 1) When the adjective follows the noun.
- 2) When the adjective is nominalized.

It should be mentioned that Lunt does not present any arguments as to *why* this is so, nor does he mention participles. Looking at other languages, we find various patterns for the same kind of expression. See example (158) below. As can be seen from table 32, among the nominalizations there is a certain amount of LF (10), and among the attributes, some SF (7).

5.4.1. Nominalized vocatives

The IS for nominalizations show almost all OLD, as is expected. All nominalized adjectives should be in the SF according to Lunt, but among the 10 LF there are 4 adjectives. Of the 14 SF, all are adjectives. For these SF, it is of interest to note that many of the same lemmas appear: *maloverъ* (5) “of little faith”, *ꝥrodъ* (4) “stupid”, *bezumъnъ* (2) “foolish”, *bui* (2) “blind”. The vocative constructions they appear in seem *quality ascribing* in nature: the speaker is ascribing a characteristic to the addressee. This seems like a more satisfying explanation for their SF use than stating that they are nominalized adjectives. There is a possibility that they are short because they correspond to anarthrous Greek constructions – but so do almost all vocatives. If the vocative expressions were referential, as in example (158), we would expect LF. Cf. the different translations of (156) and (157).

- (156) ꙗ́гла имъ. что страшиви есте **маловѣри**. (CM, Matt 8:26, 50736)
 kai legei autois, ti deoloi este, **oligopistoi**; (47772)
 Why are you afraid, O **you of little faith**?
 Pourquoi avez-vous si peur? leur dit-il. **Votre foi est bien petite!**
- (157) **безоумъни** не иже ли естъ сътворилъ вънѣштнее. ꙗ́ вънѣштнее сътвори. (CM, Luke 11:40, 40704)
 afrones, oukh ho poiēsas to exōthen kai to esōthen epoiēsen; (20951)
 You fools! Did not he who made the outside make the inside also?
 Vilken skenhelighet! Ni vet ju i alla fall att Gud har skapat både det yttre och det inre.
 Fous que vous êtes! Est-ce que celui qui a créé l'extérieur n'a pas aussi fait l'intérieur?

How are the 10 LF explained? They correspond to restrictive relative clauses. Such nominalized participles are often quantified, and most often long. An example is found in (158):

- (158) идѣте отъ мене **проклѣтии** въ огнь вѣчны... (CM, Matt 25:41, 39571)
 poreuesthe ap' emou **katēramenoi** eis to pur to aiōnion... (15992)
 “**You that are accursed**, depart from me into the eternal fire...” (relative clause)
 “Gå bort med er, **ni förbannade**, till den eviga eld...” (per. pronoun + def participle)
 “Retirez-vous de moi, **maudits**; allez dans le feu éternel...” (unmarked participle)

“You that are accursed” is rendered by a nominalized participle in OCS. The speaker knows who he is talking to, i.e. this is not a case of a generic or quantified referent¹⁰², but a specific one. 6 of the above mentioned 10 long vocatives have dependents (underlined in the example below) that further delimit or anchor the referent, and this could also trigger LF, as in (159). This very sentence is found with SF as well in Matt 25:34.

- (159) идѣте **благословѣнїи** отъ ца моѣго (CS, Vita of Basiliskos, 82309)
 Go, **you who are blessed by my father!**

¹⁰² He is addressing the people on his left as opposed to the righteous ones on his right hand that are addressed in MATT 25:34-40, and whom he continues to address in the lines that follow. The nominalization has therefore been analyzed as OLD in the IS analysis of the corresponding Greek text, where *proklēti* corresponds to the participle *kathramenoi*, which eventually refers back to *ex evōnumōn* (from the ones on the left).

Lunt's statement that nominalized vocatives display SF is therefore supported here with 75.0%, but the results are hardly statistically significant¹⁰³, and it is more interesting to look at *why* these particular adjectives show LF, notably that they are possibly quality ascribing.

5.4.2. Attributive vocatives

As for attributes to vocatives (35), the IS distribution looks very much the same, with predominantly OLD. What is unexpected here is the number of SF (7). All of these follow their heads, and we recall that Lunt argued that such post-nominal attributes are short. He does not, however, say why this is, and other explanations are also available. In Luke 9:41, the "i" in *o rode nevěryň i razvrašteně* could affect the graphic outcome of *nevěryň*, which is here erroneously written with a front *jer* and should really be *nevěryňi* or *nevěryny* in LF. Compare the sentences below :

- (160) ѿ роде невѣрънь ѿ развраштенѣ до колѣ бѣдѣ въ васъ и трѣплѣ вы. (CM, Luke 9:41, 40529)
 ō genea **apistos** kai **diestrammenē**, heōs pote esomai pros humas kai anexomai humōn; (20766)
 You **faithless** and **crooked** generation, how long will I be with you and put up with you?
- (161) ѿ роде невѣръны и развращены. до колѣ съ вами бѣдѣ (CM, Matt 17:17, 50893)
 You **faithless** and **crooked** generation, how long will I be with you?

The adjective *nevěryň* also appears in another nearly identical sentence, but without the following "i". Three of the post-positioned adjectives are *ijudeiskъ*, which has been discussed many times as it almost only shows SF. Another SF is the adjective *ijerusalimъskъ*, which also belongs in the group with "untrustworthy" use of SF/LF. The last SF is the only one actually displaying a vocative ending different from the nominative: *fariseju slepe*, "O blind Pharisee!" (Matt 23:26). This exclamation is part of a long row of vocatives in Matt 23. Most show LF, but most are also plural, which has no specific vocative form. This could be a reason for this particular instance showing SF, rather than just its position in relation to its head word. This idea is strengthened by Rondestvedt (1986, p. 125), who says that this particular SF vocative is an alternative to the LF nom. for masc. sg. vocatives.

5.4.3. Conclusion regarding vocatives

Vocatives mainly display LF since in addressing someone, the speaker will usually know who the person he is addressing is. This is confirmed by corpus data, showing that the OLD tag is the tag most frequently found on referents in the vocative. When exceptions occur, position in relation to its head word, nominalized status or influence from Greek could play some role. Another explanation would be the semantics underlying the particular lemmas or the pragmatics of the phrase. If the vocative is

¹⁰³ P-value = 0.02266, using a binomial test.

specific, referring to a specific person or group of people, LF is expected, but if it ascribes a quality to said person/group it would make sense for it to be short.

6. Concluding remarks

6.1. Summary and findings

This thesis has examined the role short and long form in adjectives and participles plays in conveying information status in texts belonging to the Old Church Slavonic canon. The main questions asked were whether SF/LF on regular attributive and nominalized adjectives and participles are true markers of definiteness or not, and if this system of explicit marking of definiteness was still in use in the OCS of the 9th to 11th century. What is here referred to as “definite” is a referent that is identifiable due to previous mention, world knowledge, deixis, association to another entity, or some other factor. In order to answer these questions, the use of SF/LF has been examined in relation to the IS of either the modified referents, or – in the case of nominalizations – the status of the adjectives and participles themselves. Attributes were first examined, sorted into IS categories, and then nominalizations. Finally, appositions as well as vocatives were quickly looked into.

Once all categories were analyzed and the results checked for statistical significance, short and long form was found to depend on the pragmatics of the phrase in which they are employed to a very large degree, confirming the traditional view that LF represents definiteness and SF indefiniteness. Previous research has suggested that LF is overused with nominalizations in OCS, particularly in the plural. We have seen that this is not the case in the data set used in this thesis.

Certain exceptions exist: occasionally, semantic factors in the underlying noun in the case of denominal adjectives decide the form of the adjective. This is notably the case with -ov and -in adjectives, which seem to behave in a determiner-genitive manner, automatically rendering their head word definite, meaning that these adjectives, regardless of strength, act as the equivalent of regular LF adjectives. This is also the case with notably -ъsk adjectives formed from unique entities (displaying mainly LF), as these adjectives will almost always anchor the referent to some well-known entity, so that even if said referent is formally new in the narrative or altogether non-referential, it appears as definite.

Overall, the feature of “anchoring” has been shown to play a large role: whenever a referent is anchored to a previously known or altogether well-known entity, it is easier to identify by the hearer, and thereby marked by LF. This is what explains a great deal of the “unexpected” LF use found in the corpus data. Generics have also been shown to make use of LF, strengthening the idea that generic referents are definite. The abundance of plurals among the LF generics is probably the reason Flier

considered LF overly frequent in plural nominalizations and plurals in general, as he only admits singular referents as generic.

As for vocatives and appositions, neither was very numerous in the narrowed down dataset, but what data there is indicates that OCS marks vocatives with LF. There may be a split between identifying vocatives and quality ascribing vocatives – explaining unexpected use of SF – similar to that for appositions. For appositions the results show that it would be beneficial to be able to tag identifying appositions (for which LF is expected, and where both the head word and the apposition refer to known entities) separately from quality ascribing appositions (where SF is expected).

SF for predicatives seems to be securely in place in these early texts, with very few occurrences of LF (which are identifying). This is understandable from a pragmatic point of view, and the variation in strength is found in attributes and appositions, as well as among nominalizations.

The table below summarizes the SF/LF for the most important IS-tags, showing that there are no statistically significant differences between the behavior of attributes and nominalizations.

Table 33 Summary of SF/LF distribution for IS-tags

| IS-tag | Attributes | | Nominalizations | | p-value* |
|--------------|------------|-------|-----------------|----------------------|----------|
| | SF | LF | SF | LF | |
| NEW | 98.6% | 1.4% | 80.0% | 20.0% ¹⁰⁴ | 0.6783 |
| NEW-anchored | 91.3% | 8.7% | 6.8% | 93.2% | 1.0000 |
| NON-SPEC | 89.8% | 10.2% | 89.8% | 10.2% | 1.0000 |
| QUANT | - | - | 7.8% | 92.2% | - |
| OLD | 4.9% | 95.1% | 6.7% | 93.3% | 0.8070 |
| ACC-GEN | 5.4% | 94.6% | 0.0% | 100.0% | 0.8923 |
| KIND | 12.5% | 87.5% | 6.9% | 93.1% | 0.8863 |

*this is the p-value for a 2x2 contingency table tested with a two-tailed Fisher's exact test, testing if the difference between the two categories is statistically significant. The contingency table contains the number of SF or LF out of the total number of occurrences for each IS-category of attributes up against the number of SF or LF for the same category of nominalizations.

As for word order, it has been found to be of little importance for these texts as it very closely mimics the word order used in the original Greek texts.

¹⁰⁴ This high percentage is misleading due to very few occurrences. See the summary in chapter 5.2.1.2.

6.2. Evaluation of the procedure

The data processing method used for this thesis proved to be more or less indispensable. It has made it possible to sort and arrange a large amount of data and make this study as quantitative as possible, while at the same time efficiently excluding particular groups of words that would not be of interest to look closer at. Earlier research has been carried out in a much more time consuming manner and without the benefit of large amount of digitalized data that can be easily processed. The use of the PROIEL corpus IS-tagging has been invaluable, as this has made it possible to not only look at individual occurrences of SF or LF, but to see their use in a larger context thanks to the tags they have been accorded by the annotators. The IS-tagging poses certain problems and certain categories of words are notoriously difficult to tag, but overall it provides very valuable insights.

6.3. Further research

As this thesis deals with a relatively great amount of categories of words, more research into each separate group could most surely reveal some interesting things. Overall, it has been felt that the IS explains the appearance of SF/LF well enough on its own to have to delve into other categories, such as syntactic function. Possibly the most interesting aspect to look at further would be the development in younger texts. Not a very great amount of texts from Codex Suprasliensis have been added to the corpus and annotated for Information Status as of yet, but they are being added now and will continue to be added in the close future. When more texts are available, comparative studies can be carried out. As it stands now, whenever CS deviates from CM, there is usually too little data to say anything conclusive regarding the significance of these differences.

Works Cited

- Archaimbault, S. (1992). L'adjectif dans la tradition grammaticale russe. *Histoire Épistémologie Langage*, 211-221.
- Birnbaum, J. D. (n.d.). *Paul the Simple, Codex Suprasliensis, 86r–88v*. Retrieved May 1, 2013, from http://clover.slavic.pitt.edu:8080/exist/paul/data/paul_main.html
- Borodič, V. V. (1963). O kategorii opredelennosti-neopredelennosti v staroslavjanskom jazyke. *Slavjanskaja Filologija* 5, pp. 162-202.
- Čaburgaev, G. (1974). *Staroslavjanskij jazyk*. Moscow: Prosvešenie.
- Dahl, Ö. (2004). Definite articles in Scandinavian: Competing grammaticalization processes in standard and non-standard varieties. In B. Kortmann, *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective (Trends in Linguistics: Studies & Monographs)* (pp. 147-180). Berlin: Mouton de Gruyter.
- Dimitrova-Vulchanova, M., & Vulchanov, V. (2009). Linearization Patterns in the Old Bulgarian Nominal Expression. In M. Dimitrova-Vulchanova, & O. M. Tomic, *Investigations in the Bulgarian and Macedonian Nominal Expression* (pp. 74- 94). Trondheim: Tapir Academic Press.
- Dost, A., & Gribanova, V. (2006). Definiteness Marking in the Bulgarian. *Proceedings of the 25th West Coast Conference on Formal Linguistics* (pp. 132-140). Somerville, MA: Cascadia Proceedings Project.
- Eckhoff, H. M. (2011, July 29). *Animacy and differential object marking in Old Church Slavonic*. Retrieved April 23, 2013, from <http://www.hf.uio.no>: <http://www.hf.uio.no/ifikk/english/research/projects/proiel/Activities/proiel/publications/anim.pdf>
- Eckhoff, H. M. (2011). *Old Russian possessive constructions*. De Gruyter.
- Flier, M. S. (1974). *Aspects of Nominal Determination in Old Church Slavic*. The Hague: Mouton & Co.
- Fraurud, K. (1990). Definiteness and the Processing of Noun Phrases in Natural Discourse. *Journal of Semantics*, 395-433.
- Greenberg, J. H. (1978). How does a language acquire gender markers? In G. J. H., C. A. Ferguson, & E. A. Moravcsik, *Universals of human language, vol. 3: Word structure* (pp. 47-82). Stanford: Stanford University Press.
- Haug, D. (n.d.). *Guidelines for annotation of givenness*. Retrieved April 24, 2013, from http://folk.uio.no/daghaug/info_guidelines.pdf
- Haug, D. T. (2010). *PROIEL Guidelines for Annotation*. Retrieved April 23, 2013, from <http://www.uio.no/>: http://folk.uio.no/daghaug/syntactic_guidelines.pdf

- Haug, D., Eckhoff, H., & Welo, E. (n.d.). The theoretical foundations of givenness. In K. Bech, & K. Eide, *Information Structure and Syntactic Change in Germanic and Romance Languages (forthcoming)*. Oslo.
- Ivanova, T. (1977). *Staroslavjanskij jazyk*. Moscow: Vyššaja škola.
- Keenan, C. (2011, June 5). Greenberg Revisited: Diachronic development of article systems & the structure of DP. *The 13th International Diachronic Generative Syntax Conference*. Philadelphia: University of Pennsylvania.
- Kurz, J. (1963). Problema člena v staroslavjanskom jazyke (konstrukcii suščestvitel'nogo s mestoimeniem ukazatel'nym i dal'nejšim atributom). In *Issledovanija po sintaksisu staroslavjanskogo jazyka* (pp. 121-182). Prague: Československaja akademija nauk.
- Le Guillou, J.-Y. (1984). *Grammaire du Vieux Bulgare (Vieux Slave)*. Paris: Jean Maisonneuve.
- Leskien, A. (1992). *Handbuch der altbulgarischen (alkirchenslavischen) Sprache*. Heidelberg: Carl Winters Universitätsbuchhandlung.
- Lunt, H. (1989). Common Slovene and Common Slavic. *Slavistična revija*, pp. 7-14.
- Lunt, H. (2001). *Old Church Slavonic Grammar*. Berlin, New York: Mouton de Gruyter.
- Lyons, C. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- Mathiassen, T. (2009). *Russisk Grammatikk*. Oslo: Universitetsforlaget.
- Meyer, C. F. (1992). *Apposition in contemporary English*. New York: Cambridge University Press.
- Napoli, M. (2009). *Aspects of definiteness in Greek*. Pisa: John Benjamins Publishing Company.
- Pelletier, F. (2010). *Kinds, Things, and Stuff - Mass Terms and Generics*. Oxford: Oxford University Press.
- Rondeštedt, K. A. (1986). *Definite/Indefinite and related pragmatic categories in Early Original Slavic*. Chicago: The University of Chicago.
- Språknämnden, S. (2005). *Språkriktighetsboken*. Svenska Språknämnden & Nordstedts Akademiska Förlag.
- Šuležkova, S. (2011). *Frazeologičeskij slovar' staroslavjanskogo jazyka*. Flinta Nauka.
- Tolstoj, N. I. (1957). Značenie kratkix i polnyx form prilagatel'nyx v staroslavjanskom jazyke (na materiale evangel'skix kodeksov). *Voprosy slavjanskogo jazykoznanij* 2, 43-122.
- Vaillant, A. (1942). L'article en vieux slave. *Revue des études slaves*, 5-12.
- Vlasto, A. (1986). *A Linguistic History of Russia to the End of the Eighteenth Century*. Oxford: Clarendon Press.
- Wade, T. (2007). *A Comprehensive Russian Grammar*. Blackwell Publishing.

Zverkovskaja, N. (1986). *Suffiksál'noe slovoobrazovanie russkich prilagatel'nych XI–XVII vv.* Moskva: Nauka.

List of tables

| | |
|--|-----|
| Table 1 Full declension of the pronoun <i>*jb</i> | 11 |
| Table 2 Short form and long form endings of adjectives and participles, soft/hard, M N..... | 12 |
| Table 3 Suffixes and their distribution in the PROIEL corpus..... | 22 |
| Table 4 The original distribution of adjectives/participles in the texts present in the corpus, prior to any exclusions..... | 27 |
| Table 5 Tags in the PROIEL corpus (from the forthcoming Haug et al., p. 7) | 28 |
| Table 6 Distribution of SF/LF across the most common IS-tags in the PROIEL corpus | 34 |
| Table 7 SF/LF distribution for the quantifiers <i>mnogъ</i> & <i>vsjěkъ</i> | 35 |
| Table 8 SF/LF distribution for the numeral <i>drugъ</i> | 35 |
| Table 9 Distribution of SF/LF for the present participle 'sy' | 36 |
| Table 10 Distribution of demonstratives cooccurring with adjectives or participles | 38 |
| Table 11 Correspondence between Greek and OCS demonstratives | 38 |
| Table 12 Error percentages for -ъsk adjectives | 48 |
| Table 13 LF/SF distribution in the NEW category for attributive adjectives and participles | 51 |
| Table 14 LF/SF distribution in the NEW-anchored category for attributive adjectives and participles | 53 |
| Table 15 Distribution of lemmas in the OLD long form category..... | 54 |
| Table 16 SF/LF distribution in the OLD category for attributive adjectives and participles | 59 |
| Table 17 LF/SF distribution in the ACC-GEN category for attributive adjectives and participles | 61 |
| Table 18 LF/SF distribution in the NON-SPEC category for attributive adjectives and participles | 68 |
| Table 19 Distribution of number and strength in the KIND category, after exclusions | 69 |
| Table 20 LF/SF distribution in the KIND category for attributive adjectives and participles | 70 |
| Table 21 Word order in Greek and OCS | 72 |
| Table 22 Summary of the distribution of number for adjectival and participial nominalizations | 74 |
| Table 23 LF/SF distribution in the NEW category for nominalizations | 77 |
| Table 24 LF/SF distribution in the NEW-anchored category for nominalizations..... | 79 |
| Table 25 SF/LF distribution in the OLD category for nominalizations..... | 82 |
| Table 26 LF/SF distribution in the ACC category for nominalizations | 84 |
| Table 27 LF/SF distribution in the NON-SPEC category for nominalizations..... | 89 |
| Table 28 Number and strength for nominalized generics..... | 89 |
| Table 29 LF/SF distribution in the KIND category for nominalizations | 91 |
| Table 30 Distribution of SF/LF among participles and adjectives in the QUANT category | 92 |
| Table 31 LF/SF distribution in the QUANT category for nominalizations | 93 |
| Table 32 Vocatives and the distribution of short/long form..... | 100 |
| Table 33 Summary of SF/LF distribution for IS-tags..... | 104 |

Appendix 1 - Change of IS-tag category

In order to access these sentences, go to http://foni.uio.no:3000/users/sign_up to sign up (for free) for access to the PROIEL corpus. For sentence IDs, use the link <http://foni.uio.no:3000/sentences/XXXXX>. For token IDs (from which the sentence is then easily accessed), use <http://foni.uio.no:3000/tokens/XXXXXX>. The search field in the PROIEL interface can also be used to access various IDs.

The sentences are followed by the sentence-ID, and the second (in brackets) is the unique token ID.

Attributes

KIND THAT CAN BE NON-SPEC:

MATT 9.17 -- нъ вино **ново** въ мѣхы новы вѣливажъ. ꙗко оное съблюдетъ сѧ.-- 38566 (566740)
LUKE 5.38 -- нъ вино **ново** въ мѣхы новы вѣливати. ꙗко оное съблюдетъ сѧ. -- 40218 (587455)
MATT 7.17 -- тако всѣко дрѣво добро. плоды добры творить. а **зѣло дрѣво плоды зѣлы** творить. -- 38482 (565483 & 565486)
LUKE 23.31 -- зане аште въ **сыръ дрѣвъѣ** си творятъ. въ соусть что бждетъ -- 41494 (603482)
JOHN 12.24 -- аште зръно пшенично не оумьретъ падъ въ земли. то едино прѣбываать -- 52033 (615159)
LUKE 12.28 -- аште ли же трѣвъ дньесъ на селѣ **сжщж**. и оутрѣ въ пешть въметаемоу -- 40767 (594571)
LUKE 11.44 -- Горе вамъ кѣнижъници и фарисѣи. ꙗко лицемѣри. ꙗко де есте ѣко гробы не **вѣдоми**. ꙗко члвци ходащеи врѣхоу не вѣдате. -- 40709 (593830)
LUKE 11.46 -- ѣко накладеате на члвкы брѣмена не оудобѣ **носима**. а сами ни единѣмъ же прѣстомъ вашимъ прикасаате сѧ брѣменехъ. -- 40713 (851425)
JOHN 4.23 -- нъ градетъ година и нынѣ есть. егда **истиньни поклоньници** поклонатъ сѧ бгцю. дхомъ и истинюмъ. -- 41796 (607167)

NON-SPEC THAT CAN BE KIND (OR NON-SPEC-ANCHORED)

MARK 4.32 -- ꙗко бждетъ боле всѣхъ зелии. и творить вѣтви велии. ѣко мошти подъ сѣниѣ его птицамъ небскимъ витати. -- 36547 (541102)
MATT 13.4 -- ꙗко сѣяшюмоу. ова оубо падж при пжти. ꙗко придж птица небскимъ и позобаша ѣ.-- 38776 (569430)
LUKE 8.53 -- егда сѣаше ово паде при пжти и поплърано бысть. ꙗко птица небскимъ позобаша е.-- 40385 (589617)
LUKE 13.19 -- и въздрате и бысть въ дрѣво велие. и птица небскимъ вселиша сѧ въ вѣтви его. - 58154 (595597)
Vita of Paul and Juliana -- 'Аурилиянъ же оубоявъ сѧ народа ꙗко крамолж въздвиг'нжъ. дасть о ѣею отъвѣтъ повелѣвъ главѣ ѣма отъсѣшти. ꙗко телесъ ѣею поврѣшти п'сомъ ꙗко звѣремъ. ꙗко птицамъ небскимъ. -- 81248 (1155131)
MATT 12.43 -- егда же **нечисты дхъ** изидеть отъ члвка. прѣходитъ скозъ бездѣнаа мѣста. ишта покоѣ и не обрѣтаатъ. -- 38758 (569209)
Vita of Paul and Juliana -- блждѣникъ сы ѣзлиха паче всѣхъ чловѣкъ. живжштиихъ. вида чловѣчскы **добрыя жены** и дѣштери. -- 81042 (1154064)

NON-SPEC THAT CAN BE ACC-GEN

MARK 3.29 -- а иже власвимісаатъ на сѣты дхъ. не имать отъпоуштениѣ въ вѣкъ. нъ повиненъ есть

въчъноумоу сждоу --36498 (540418)

JOHN 4.36 -- жънаи мѣздѣхъ приемлетъ. ꙗко собираѣхъ плодъ въ **животъ въчъныи**. да и сѣмѣи въкоупѣ радуоуотъ сѣ и жънаи. -- 41820 (607398)

JOHN 10.28 -- азъ **животъ въчъныи** даждѣ имѣ. ꙗко не погыбнѣтъ въ вѣкъ. ꙗко не въсхытитъ ихъ никтоже отъ рѣкъ моеѣ. -- 42362 (613431)

JOHN 12.25 -- любѣи дѣждѣ своѣхъ погубитъ ѿ. ꙗко ненавиждѣи дѣждѣ своеѣ. въ мирѣ сѣмѣ. въ **животѣ въчънѣемѣ** съхранитъ ѿ. -- 42504 (615194)

JOHN 17.2 -- ѿкоже далъ емоу еси власть въсѣкоѣ плѣти. да въсѣко еже еси емоу далъ дасть имѣ **живота въчънааго**. -- 42762 (618348)

Vita of Paul and Juliana -- не примышляѣи ми **сѣмрѣти въчъныи** лишити мѣ хотѣ славы хѣсовѣ. и цѣсарѣства небесънааго. ѿкоже ты штоуждѣ ѿси. -- 80882 (1153133)

NEW THAT CAN BE ACC-SIT OR ACC-GEN (OR POSSIBLY ANCHORED)

MARK 15.38 -- **опона цѣрквнаѣ** раздѣра сѣ на дѣвое сѣ выше до ниже. --37331 (550791)

LUKE 23.45 -- и помрѣче слѣнѣце. ꙗко **катапетазма цѣрквнаѣ** раздѣра сѣ на дѣвое. -- 41518 (603729)

MATT 9.15 -- рече имѣ ѿсѣ. еда могутъ **снѣве брачънии** плакати сѣ. донѣдеже сѣ ними естѣ женихъ. -- 50763 (566660)

LUKE 5.34 -- онъ же рече къ нимѣ. еда можете **снѣи брачъныи** донѣдеже женихъ естѣ сѣ ними сѣтворити постити сѣ. 51302 (587362)

JOHN 15.11 -- азъ есмѣ **лоза истинънаѣ**. ꙗко бѣтѣцъ мои дѣлатель естѣ. -- 42667 (617066)

LUKE 22.66 -- ꙗко быстѣ дѣнь сѣбѣраша сѣ **старѣци людѣсци**. -- 41437 (602863)

MATT 17.25 -- **цѣсри земѣсци** отъ кѣхъ приемлѣтъ дани. ли кинсѣ. -- 39075 (572852)

MATT 18.68 -- да обѣсѣтъ жрѣновъ на вѣи его осѣльскѣи. ꙗко потопѣтъ ꙗко въ **пѣчинѣ морѣстѣи**. -- 39088 (573036)

MATT 23.33 -- змиѣи ищѣдѣ ехидънова како оубѣжите отъ **сѣда леонѣскааго**. -- 39444 (577245)

MARK 14.62 -- оузѣрите сѣна цѣлскааго. о дѣснѣхъ сѣдашта силѣи. ꙗко градѣшта сѣ **облакѣи ѿбсскѣими**. -- 37256 (549994)

MATT 24.29 -- абие же по скрѣби дѣнии тѣхъ. слѣнѣце мрѣкнетѣ. и лоуна не дасть свѣта своего. ꙗко звѣзды сѣпадѣтъ сѣ ѿбсе. ꙗко **силѣи небескѣи** двигнѣтъ сѣ. -- 39487 (577869)

LUKE 21.26 -- издѣхѣахѣстѣмѣ цѣлкѣмѣ отъ страха. и чааниѣ градѣштихъ на въселенѣхъ. **силѣи** бо **ѿбсскѣи** подвигнѣтъ сѣ. -- 51680 (601527)

MATT 26.64 -- отъ сѣлѣ оузѣрите сѣна цѣлвѣчскааго. сѣдашта о дѣснѣхъ силѣи. ꙗко градѣшѣ на **облацѣхъ ѿбсскѣихъ** -- 39690 (851894)

JOHN 15.11 -- азъ есмѣ **лоза истинънаѣ**. ꙗко бѣтѣцъ мои дѣлатель естѣ. -- 42667 (617066)

MATT 10.15 -- отѣрадѣнѣ бѣдетѣ земѣи содомѣсцѣи гоморѣсцѣи. въ **дѣнь сѣдѣныи**. неже градуо томоу -- 50783 (567378)

Vita of Paul and Juliana -- ꙗко потомѣ пригваждаѣтъ сѣ на **дрѣвѣ крѣстѣнѣмѣ**. страждѣ мирѣскааго ради сѣпасѣний. сѣнабѣдѣти мирѣ погыбѣшѣ. по прѣлѣштѣниѣ дѣяволѣ. -- 81163 (1154683)

Vita of Paul and Juliana -- посѣла бо ѿгѣла своѣго сѣ **высотѣ небесѣсскѣихъ** съхранити ѿ -- 81752 (1153600)

LUKE 6.11 -- Быстѣ же въ **сѣботѣхъ вѣторопрѣвѣхъ**. -- 40221 (587484)

MATT 26.33 -- тѣгда сѣбѣраша сѣ архiereи и кѣнижѣници и **старѣци людѣстѣи**. на дворѣ архiereовѣ. нарицаемааго каиѣфа. -- 39580 (579170)

NEW-ANCHORED THAT CAN BE NEW OR NON-SPEC:

MARK 5.30 -- абѣе ѿсѣ оштѣшѣ въ сѣбѣ **силѣи ишедѣшѣ** отъ него. обрѣштѣ сѣ въ народѣ гѣлаше. -- 36595 (541790)

MATT 13.47 -- Пѣкѣи подобѣно естѣ цѣсрѣстѣиѣ ѿбсское **неводоу вѣврѣженоу** въ морѣ. ꙗко отъ въсѣкого рода избѣравѣшѣ. -- 38841 (570285)

MARK 1.68 -- бѣ же иоанѣ обѣлѣченѣ власѣи вѣлѣбѣждѣи. ꙗко **поѣсѣ оусниѣнѣ** о чрѣслѣхъ его. ꙗко ѣдѣ акриди и медѣ дивии. -- 36365 (538625)

JOHN 5.55 -- Бѣ же тоу единѣ **цѣлвѣкѣ**. ѿ. ꙗко осмѣ лѣтъ **имѣи** въ недѣжѣ своеѣ. -- 41853 (607866)

MARK 14:18 -- амин ꙗкоже вамъ. ꙗко единъ отъ васъ прѣдасть мѧ. **ѣды** съ мноуж -- 37182 (549189)
 MATT 13.24 -- Оуподоби сѧ црствие небеское. члвкоу съвъшоу добро съма на селѣ своемъ. -- 38802 (569823), (39187 (574331) is very similar)
 MATT 13.31 подобно естъ црствие ꙗбское. зрюноу гороушню. еже въземъ чловѣкъ въ сѧ на селѣ своемъ.-- 38816 (569973), (40836 (595576) is identic to this sentence).
 LUKE 1.27 -- къ дѣвѣ обржченѣ мжеви. емоуже има иосифъ. отъ домоу давдва і има дѣвѣ мариѣ. 39947 (583391)
 LUKE 18.91 -- рече же и къ единѣмъ надѣжштемъ сѧ на сѧ. ꙗко сжтъ праведѣници. и оуничѣжающемъ прочѧмъ притѣчѧ сѧ -- 41085 (598765)
 39349 (576121), 40156 (586677), 36466 (539926), 58233 (598635), 39799 (581484), 41540 (604133), 36421 (539358), 38838 (570220) – See Appendix 2.

OLD THAT CAN BE NON-SPEC:

LUKE 19.71 -- і видѣвъше вси рѣптаахъ ꙗкоже. ꙗко къ грѣшнѣмъ мжю въниде витатѣ. -- 41149 (599497)
 MARK 7.8 -- оставше бо заповѣдъ бжѣ. дрѣжите прѣданиѣ члвчска. -- 36702 (543321)
 MATT 13.50 -- і въврѣжѣтъ ѧ въ пещѣ огнѣнѣ. -- 38845 (570337)
 MATT 13.23 -- а сѧное на добрѣ зѣми. съ естъ слышѧи слово и разоумѣваѧ е. -- 38800 (569781), (40399 (589807) and 36529 (540860) feature the same expression)
 MARK 3.33 -- ꙗко члкоу имжштоумоу соухѣ ржѣ -- 50264 (539947)
 Vita of Paul and Juliana -- никѣтоже бо можааше свойма очима ба бесплѣтна видѣти. -- 81154 (1154630)

ACC-GEN THAT CAN BE NON-SPEC:

LUKE 7.25 -- се иже въ одежди славнѣ і пишти сжште. въ црхѣ сжтъ. -- 40334 (589014)
 LUKE 7.32 -- Подобни сжтъ отрочиштемъ сѣдаштемъ на трѣжиштихѣ. -- 40344 (589130)
 MATT 13.42 і въврѣжѣтъ ѧ въ пещѣ огнѣнѣ. -- 38834 (570186)
 JOHN 5.35 -- онъ бѣ свѣтильникъ гора и свѣта. -- 51856 (608458)

Nominalizations

NEW THAT CAN BE KIND:

MATT 21.14 -- і пристѣпиша къ нему хроми и слѣпи въ цркве и исцѣли ѧ. -- 39263 (575235)

NEW THAT CAN BE QUANT:

MARK 6.31 -- бѣахѣ. бо приходаштии. і оходаштеи мнози -- 36659 (542678)

NEW THAT CAN BE ACC-INF:

MARK 16.19 -- ꙗкоже ѡ по ꙗлании его къ нимъ. възнесе сѧ на небо. і сѣде о деснѣхъ ба. -- 37368 (551323), 41442 (800334) and 39690 (851888) feature the same expression.

NEW-ANCHORED THAT CAN BE ACC-INF:

MARK 5.14 -- і пасхѣти свинѣмъ бѣжашѧ. і възвѣстиша въ градѣ и на селѣхѣ. -- 36580 (541492)
 MARK 10.37 -- даждѣ нама да единъ о деснѣхъ тебе. и единъ о шюхъ тебе сѣдевъ въ славѣ твоѣ -- 50462 (546298), 50579 (550614) and 81662 (1154757) feature the same expression.

NON-SPEC THAT CAN BE OLD:

LUKE 5.39 -- і никѣтоже пивѣ ветѣха абие хоштетѣ новоуоумоу. -- 40219 (587472)

NON-SPEC THAT CAN BE QUANT:

LUKE 4.40 -- заходяшту же слънѣцоу. вѣси елико имѣахъ **болащамъ** неджгы различъны. привождаяхъ ѿ къ нему. -- 51280 (586542)

LUKE 10.16 -- **Слоушаѿи** васъ мене слоушааѿ. ꙗко отъмѣтаѿи сѧ васъ мене сѧ отъмѣтааѿ. -- 40591 (592330)

NON-SPEC THAT CAN BE KIND:

MARK 6.55 -- ꙗко прѣтѣша въсѣ странѣ тѣхъ. ꙗко начаса приносити на одрѣхъ **болащамъ**. идеже слышаахъ ꙗко тоу естъ. -- 36693 (543113)

MATT 10.8 -- **болащамъ** цѣлите. **мрътвымъ** въскръшаите. -- 50776 (567230 & 567233)

MATT 10.8 -- **прокаженнымъ** очиштаите. -- 50778 (567236),

LUKE 1.53 -- **алчашамъ** исплѣни блага. ꙗко богаташамъ сѧ отъпоусти тѣшѧ. -- 51237 (583798)

MATT 7.67 -- Не дадите **ѣтаго** псомъ. ни помѣтаите бисѣрь вашихъ прѣдъ свинѣми. -- 38464 (565266)

MATT 11.57 -- **слѣпии** прозираѿтъ. ꙗко **хромии** ходатъ. **прокажени** очищаѿтъ сѧ и **глоусии** слышатъ. **мрътвии** вѣстаѿтъ. ꙗко **ништи** благовѣстоуѿтъ. -- 38658 (567949, 567953, 567956, 567960, 567963, 567967)

LUKE 7.21 -- шѣдѣша възвѣстѣи иоанови. ѣже видѣста и слышаста. ꙗко **слѣпии** прозираѿтъ.

хромии ходатъ. **прокажени** очищаѿтъ сѧ. **глоусии** слышатъ **мрътвии** вѣстаѿтъ. **ништи** благовѣстоуѿтъ. -- 40327 (588947, 588950, 588953, 588957, 588959, 588962)

LUKE 10.9 -- ꙗко цѣлите **неджжънымъ** иже сѣтъ въ немъ. ꙗко цѣлите имъ. -- 40581 (592203)

MATT 12.35 -- Добры члвкъ отъ добраго съкровишта износитъ **добраѣ** ꙗко злы члвкъ отъ злаго съкровишта износитъ **злаѣ**. -- 38749 (569063, 569071)

MATT 19.21 -- аште хоцещи съврѣшенъ быти. иди продаждъ имѣние твое. ꙗко даждъ **ништиимъ**. -- 50931 (574110)

MATT 26.9 -- можаше бо се муро продано быти на мнозѣ. ꙗко дано быти **нищимъ**. -- 39586 (579254)

LUKE 14.13 -- нъ егда твориши пиръ. зови **нишѧ** маломошти. хромы слѣпы. -- 40882 (596180)

LUKE 11.21 -- егда **крѣпкы** оуоржжъ сѧ хранитъ свои дворъ. въ мирѣ сѣтъ имѣниѣ его. -- 40675 (593354)

LUKE 1.17 -- ꙗко тѣ прѣдъидетъ прѣдъ нимъ дѣхомъ и силоу илииноу. обратити срѣдѣца отъцемъ на чада. ꙗко **противнымъ** въ мудрость праведныхъ. оуготовати ꙗви люди съврѣшени. -- 39934 (583191)

MATT 15.31 -- ꙗко народу дивити сѧ. выдаще нѣмы ꙗвѣша. **бѣдънымъ** съдравы. ꙗко **хромы** ходаша. **слѣпы** выдаша. -- 38969 (571629, 571626, 571622)

LUKE 16.12 -- ꙗко аште въ **тоуждемъ** вѣрѣни не бысте. ваше къто вамъ дасть. -- 40987 (597524)

LUKE 1.51 -- сътвори дрѣжавъ мышѣцежъ своежъ. расташи **грѣды** мыслижъ брѣдѣ ихъ. -- 51235 (583784)

LUKE 1.52 -- низъложи **силънымъ** съ прѣстолѣ. ꙗко възнесе **сѣмъренымъ**. -- 51236 (583790, 583796)

MARK 3.27 -- никтоже не можетъ съсѣдъ **крѣпкааго** въшедъ въ домъ его расхѣтити. аште не прѣжде крѣпкааго съважетъ. ꙗко тѣгда домъ его расхѣтитъ. -- 36496 (540364), (38742 (568928) features the same sentence.

NON-SPEC THAT CAN BE ACC-INF:

LUKE 3.13 -- ничтоже боле **повелѣнааго** вамъ творите. -- 40081 (585398)

OLD THAT CAN BE KIND:

MARK 9.10 -- ꙗко оудрѣжаша слово въ себѣ. сътазажшѣ сѧ что естъ еже из **мрътвыхъ** въскръснѣти. -- 36818 (544774), (51559 (597936), 42482 (614911) & 42491 (615051) feature the same sentence)

MARK 12.27 -- нѣстъ богъ **мрътвыхъ**. нъ бѣ живыхъ. -- 37073 (547773), (41275 (600904) features the same sentence).

MATT 20.25 -- вѣсте ѣко кѣнази ѡзкъ оустоужть имъ. и велиции обладажъ **малыми**. -- 39230 (574815)

KIND THAT CAN BE QUANT:

LUKE 16.10 -- **Вѣрны** въ **малъ** и въ мнозѣ вѣренъ есть. | **неправдыны** въ **малъ** и въ мнозѣ неправденъ есть. -- 40985 (597488, 597498, 597496, 597486)

KIND THAT CAN BE NONSPEC:

MARK 10.25 -- оудобѣе есть вельбждоу сквозѣ игълинѣ оуши проити. неже **богатоу** въ цѣрствие бжие вѣннити. -- 36923 (546029), (39175 (574173) features the same sentence)

MATT 19.23 -- амин ѣлж вамъ. ѣко не оудобѣ вѣнидеть **богаты** въ цѣрствие ѣбское -- 39174 (574153)

QUANT THAT CAN BE KIND:

MATT 9.12 -- не трѣбоужъ **сѣдрави** врача. нѣ болащеи. -- 50759 (566598)

Appendix 2 – Additional data

Examples from chapter 5.1.2.3, Attributive NEW-anchored that are erroneously anchored for various reasons

- (1) Въшедъ же църъ видѣтъ възлежащихъ. видѣ тоу ѿка не **облъчена** въ одѣание брачъное (ACC-GEN) (CM, Matt 22:11, 39349)
eiselthōn de ho Basileus theasasthai tous anakeimenous eiden ekei anthrōpon ouk **endedumenon** enduma gamou (15768)
But when the king came in to look at the guests, he saw there a man **who had** no wedding garment.
Comment: The anchoring factor is the adjective “wedding-”, which is denominal and tagged as generally accessible.
- (2) ꙗ тѣ бѣ стоѣ ѿ сѣ при езерѣ генисаретѣсѣ. ꙗ видѣвъ двѣ корабица (NEW) **стоѣшта** при езерѣ (OLD) (CM, Luke 5:1, 40156)
egeneto de en tō ton okhlon epikeisthai autō kai akouein ton lomon tou theou, kai autos ēn hestōs para tēn limēn Gennēsaret kai iden due ploiaria **hestōta** para tēn limēn (20378)
On one occasion, while the crowd was pressing in on him to hear the word of God, he was standing by the lake of Gennesaret, and he saw two boats [**standing**] by the lake...
Comment: The anchoring factor is the lake, but the lake does not make the boats identifiable or known in any way.
- (3) ꙗ вниде пакы въ сѣньмиште. ꙗ бѣ тоу ѿлвкъ (NEW) соухъ ржжъ (ACC-INF) **имы** (CM, Mark 3:1, 36466)
kai eisēlthen palin eis sunagōgēn, kai ēn ekei anthrōpos exērammenēn **ekhōn** tēn kheira (6610)
Again he entered the synagogue, and a man was there **with [having]** a withered hand.
Comment: The anchoring factor is the inferable hand, but in this sentence it could just as well be non-referential. This particular man and his hand are discussed in chapter 5.1.3.2 as well.
- (4) Сѣди (NEW) бѣ единъ. въ единомъ градѣ. ба (ACC-GEN) не **боѣ** сѣ и ѿкъ не срамлѣ сѣ. (CM, Luke 18:2, 58233)
kritēs tis ēn en tini polei ton theon mē **foboumenos** kai anthrōpon mē entrepomenos (21336)
In a certain city there was a judge who neither **feared** God nor respected man.
Comment: The anchoring factor is God, who is only used in a description of the judge’s character.
- (5) Поздѣ же бѣвъшю. приде ѿлкъ (NEW) **богатъ** отъ ариматѣѣ. именемъ (ACC-INF) иосифъ. ѿже(OLD) и оучи сѣ оу ѿ сѣ (CM, Matt 27:57, 39799)
opsias de genomenēs ēlthen anthrōpos **plousios** apo Arimatheias, to onoma lōsēf, hos kai autos emathēteuthē tō lēsou (16223)
When it was evening, there came a rich man from Arimathea, named Joseph, who also was a disciple of Jesus.
Comment: The anchoring factor is either “Jesus” or the subsequent mention of the man (autos in Greek, left out in OCS).
- (6) ꙗ се двѣ отъ нихъ бѣсте идѣшта въ тѣжде день. въ весь (NEW) **отъстоѣштѣ** стадии шесть десѣтъ отъ ѿма (OLD). еиже имѣ емаоусъ (CM, Luke 24:13, 41540)
kai idou duo ex autōn en autē tē hēmera ēsan porevomenoi eis kōmēn **apekhousan** studios exēkonta apo Ierousalēm, hē onoma Emmaous (21818)
That very day two of them were going to a village named Emmaus, [**standing**] about seven miles[a] from Jerusalem.
Comment: The anchoring factor is Jerusalem, which only indicates the location of this village that is being introduced.

- (7) ꙗ приджъ къ немуѡ носаште ослабленъ (NEW) жилами (ACC-INF). **носимъ** четырьми. (CM, Mark 2:3, 36421)
 kai erkhontai ferontes pros auton paralutikon **airomenon** hupo tessarōn (6563)
 And they came, bringing to him a paralytic **carried** by four men.
Comment: OCS literally: "carrying a weakened-in-the-limbs (person)", where the anchoring factor is "limbs" (ACC-INF). In this sentence, the OCS text has it's own IS tagging, separate from the Greek.
- (8) пакы подобно естъ црствие ѿбское. съкровищу (NEW) **съкръвеноу**. на селѣ (ACC-INF). еже обрѣтъ члвкъ съкры. (CM, Matt 13:44, 38838)
 Homoia estin hē Basileia tōn ouranōn thēsaurō **kekrummenō** en tō agrō, hon heurōn anthrōpos ekrupsen (47861)
 The kingdom of heaven is like treasure **hidden** in a field, which a man found and covered up.
Comment: The anchoring factor is the field, which obviously does not render the treasure definite.

Examples from chapter 5.2.1.3, Nominalized NEW-anchored that are erroneously anchored for various reasons

- (1) даждъ нама да единъ о деснѣжъ тебе. и единъ о **шюжъ** тебе сѣдевъ въ славѣ твоѣи (CM, Mark 10:37, 50462)
 dos hēmin hina eis sou ek dexiōn kai eis sou ex **apisterōn** kathisōmen en tē doxē sou (47336)
 And they said to him, "Grant us to sit, one at your right hand and one at your **left**, in your glory."
Comment: This sentence occurs twice, and should be tagged ACC-INF. The anchoring factor is the personal pronoun tebe.
- (2) ꙗ **пашшти** свиниѡ бѣжаша. ꙗ възвѣстиша въ градъ и на селѣхъ. (CM, Mark 5:14, 36580)
 kai **hoi boskontes** autous efugon kai apēngeilan eis tēn polin kai eis tous agrous (6728)
The herdsmen fled and told it in the city and in the country. And people came to see what it was that had happened.
Comment: This sentence should probably be tagged ACC-INF (where LF would still be expected), inferable from a previously mentioned herd of pigs. It is anchored in svinię in this sentence.
- (3) ꙗ глаголю вамъ. ѿко цркви. ꙗ **болеи** естъ съде (CM, Matt 12:6, 38704)
 lego de humin hoti tou hierou **meizon** estin hōde (15120)
 I tell you, something **greater** than the temple is here.
Comment: This is a comparison, where «the temple» anchors the nominalization, which features a lemma for which the SF/LF distinction is difficult to make anyway.
- (4) ...ꙗ се **болеи** ионъ съде. (CM, Matt 12:41, 38757)
 ...kai idou, pleion lōna hōde.
 ...and behold, **something greater** than Jonah is here.
Comment: Same explanation as for the example above.
- (5) Бѣ же единъ **бола** лазарь. отъ витаниѡ градъца мариина. ꙗ марты сестры ѡ. (CM, John 11:1, 42382)
 ēn de tis **asthenōn**, Lazaros apo Bēthanias, ek tēs kōmēs tēs Marias kai Marthas tēs adelfēs autēs (22711)
 Now a certain man **was ill**, Lazarus of Bethany, of the village of Mary and her sister Martha.
Comment: This is a clearly specific, indefinite expression where the referent is further identified by his provenance in an apposition. An OLD element in the apposition erroneously anchors the head.
- (6) нъ чесо изидете видѣтъ. ꙗрк ли еи ꙗ глаголю вамъ и **лише** ꙗрка. (CM, Luke 7:26, 51337)
 alla ti exelēluthate idein; profētēn; nai, legō humin, kai **perissoteron** profētou. (57029)

What then did you go out to see? A prophet? Yes, I tell you, and **more** than a prophet.
Comment: In “more than a prophet”, “prophet” erroneously anchors “more”.

- (7) і придѣхъ къ нему носаше **ослабленъ** жилами. носимъ четырьми. (CM, Mark 2:3, 36421)
 kai erkhontai ferontes pros auton **paralutikon** airomenon hupo tessarōn (6563)
 And they came, bringing to him a **paralytic** carried by four men.
Comment: OCS literally: “carrying a weakened-in-the-limbs (person)”, where the anchoring factor is “limbs” (ACC-INF). In this sentence, the OCS text has it’s own IS tagging, separate from the Greek. This example is also found in the section above (no 7).

Examples from chapter 5.2.2.2, OLD SF that should display LF

- (1) видѣвъше же оубо клеверѣти...**бывъша** съжалиша си сѣло. і пришедъше съказаха гноу своему въсѣ бывъшаа. (CM, Matt 18:31, 39124)
 idontes oun hoi sundouloi autou **ta genomena** elupēthēsan sfodra, kai elthontes diesafēsan tō kuriō heautōn panta ta genomena (15541)
 When his fellow servants saw **what had taken place**, they were greatly distressed, and they went and reported to their master all that had taken place.
- (2) тъгда идетъ и поиметь съ собою семь инѣхъ дѣхъ. люштьшь себе. і въшедъше живѣтъ тоу. і бѣдѣтъ **послѣднѣ** члѣва того горѣша прѣвѣхъ. (CM, Matt 12:45, 38762)
 tote porevetai kai paralambanei meth’ heautou hepta hetera pneumatata ponērotera heautou, kai eiselthonta katoikei ekei, kai ginetai **ta eskhata** tou anthrōpou ekeinou kheironta tōn prōtōn (15181)
 Then it goes and brings with it seven other spirits more evil than itself, and they enter and dwell there, and **the last state** of that person is worse than the first. So also will it be with this evil generation.
- (3) **гряды** съ выше надъ въсѣми естъ. (CM, John 3:31, 41749)
 ho anōthen **erxomenos** epanō pantōn estin (22044)
He who comes from above is above all.
- (4) отъвѣштавъ же симонъ рече непѣштѣхъ ѣко емоуже **ваште** отъда (CM, Luke 7:43, 40364)
 apokritheis Simōn eipen hupolambanō hoti hō **to pleion** exarisato (20593)
 Simon answered, “The one, I suppose, for whom he cancelled **the larger debt**.”
- (5) **боуи** бо приемъша свѣтильницы свои. не възаша съ собою олеа. (CM, Matt 25:3, 39513)
 hai gar **mōrai** labousai tas lampadas ouk elabon meth’ heautōn elaion (15933)
 For when **the foolish** took their lamps, they took no oil with them.
- (6) **боуи** рѣша мѣдрѣимъ дадите намъ отъ олѣа вашего. ѣко свѣтильници наши оугасаѣтъ. (CM, Matt 25:8, 39519)
 hai de **mōrai** tais fronimois eipan dote hēmin ek tou elaiou humōn, hoti hai lampades hēmōn sbennuntai (15939)
 And **the foolish** said to the wise, ‘Give us some of your oil, for our lamps are going out.’

Division of adjectives formed with the -ьsk suffix from chapter 5.1.1.

Uniques: *nebesьskъ, mirьskъ, cesarьskъ, človečьskъ, zemьskъ, božьskъ, geonьskъ.*

Miscellaneous: *ljudьskъ, morьskъ, možьskъ, proročьskъ, žitiьskъ, besovьskъ, detьskъ, ijereiskъ, osvьskъ, plъtьskъ, psalъmьskъ, pistynьskъ.*

Peoples & Places: *ijudeiskъ, eleonьskъ, galileiskъ, fariseiskъ, tyrьskъ, gadarinьskъ, genisaretьskъ, elinьskъ, ijerusalimьskъ, iordъdanьskъ, sodomьskъ, siluamьskъ, asiiskъ, dalъmanufanьskъ, dekaropolьskъ, egypьtьskъ, gerьgesinьskъ, južьskъ, kedrьskъ, krъstijanьskъ, kyrineiskъ, magdalanьskъ, nazaretьskъ, polьskъ, samarьskъ, samarenьskъ, tiverijadьskъ, xanaeiskъ.*

Abstract/Non-abstract division from chapter 5.2.4.2.

Abstract: *zъlъ* - 5, *dobrъ* - 4, *lichъ* - 4, *tainъ* - 4, *blagъ* - 2, *dostoinъ* - 2, *bogatъ* - 1, *gorii* - 1, *novъ* - 1, *съмрътънъ* - 1

Physical: *malъ* - 2, *kreпъ* - 1, *nemъ* - 1

Other: *съnedънъ* - 2, *mekъkъ* - 1, *vetъchъ* - 1